EFFECTOR BOTTLENECK: MICROBIAL REPROGRAMMING OF HUMAN NEUTROPHIL TRANSCRIPTION BY *ANAPLASMA PHAGOCYTOPHILUM*

by

Sara H Sinclair

A dissertation submitted to the Johns Hopkins University in conformity with the requirements of the degree of Doctor of Philosophy.

Baltimore, MD
March 2015

Abstract

Anaplasma phagocytophilum, the causative agent of human granulocytic anaplasmosis, is an obligate intracellular bacterium that infects human neutrophils. Once infected, the neutrophil shows marked functional aberrations, notably reduced antimicrobial activity, delayed apoptosis, and increased proinflammatory responses. The altered functions allow for bacterial propagation and spread to new host cells. Many of the observed phenotypic changes of the neutrophil are a result of transcriptional alterations. Furthermore, AnkA, a bacterial-derived effector localizes to the nucleus where it directly binds DNA, such as at the promoter of *CYBB* which encodes the NADPH oxidase component $gp91^{phox}$. Binding of AnkA at *CYBB* results in downregulated transcription. Furthermore, histone deacetylase 1 (HDAC) activity is important for bacterial propagation and downregulation of multiple host defense genes. While the lab previously focused on characterizing the interaction between AnkA, *CYBB*, and HDAC1, we felt the data strongly suggested that *A. phagocytophilum* must have additional transcription-regulating effectors (nucleomodulins) to act on a global scale and to explain the magnitude of phenotypic changes observed within the neutrophil during the course of infection. A bioinformatic screen of the *A. phagocytophilum* genome for genes which could encode putative eukaryotic nuclear localization signals, and mass spectrometry of proteins in the infected cell nucleus yielded nearly 50 candidate nucleomodulins. Tracking of cellular localization with GFP-fusion peptides confirmed nuclear localization of seven genes, one of which (APH_0455) is also a type IV secretion substrate. Close investigation of microarray data gathered in 2005 comparing *A. phagocytophilum*-infected and uninfected neutrophils showed that large chromosomal

territories were coordinately regulated. In addition, infected neutrophils were characterized by genome hypermethylation. Together these data further suggested that *A. phagocytophilum* infection influences transcriptional programs on a global scale, rather than relying solely on *cis* interactions and that AnkA likely is not the sole nucleomodulin at work.

Acknowledgements

I would like to thank my mentor Dr. J Stephen Dumler for always permitting me to work independently and being supportive of my ideas. His guidance and encouragement allowed me to grow into the scientist I am today. I needed to learn for myself when I had pursued something too long, and how to move on to the next questions. I would also like to thank the following current and former lab members for their continuous support both in and out of the lab: Kristen Bankert, Marguerite Lichay, Dennis Grab, Kyoung-Sung Choi, Diana Scorpio, Cindy Chen, Emily Clemens, Olga Nikolskaia, Lucy Hritzo, Carlos Borroto, and Jose Carlos Garcia-Garcia. I could not have asked for a better group of individuals to work with, and I have made life-long friends in the process.

Additionally, I would like to thank my committee members Drs. Frank DeLeo, Stephen Baylin, and Srinivasan Yegnasubramanian for their insight and thoughtful criticism.

Thank you to my family for encouraging me to leave Idaho and pursue graduate school. Without their love and support I would not have pursued research or a PhD. I would also like to thank my husband Dylan for moving across the country to support me and push me to work hard.

Lastly, I would like to thank my friends who have been like family to me over the last five years. I know that I have made life-long friendships and connections and I would not be finishing now without their help. Between my friends and the lab, my PhD experience was both an enjoyable and fruitful experience.

Table of Contents

List of Figures:

Chapter 1: The Effector Bottleneck: Chromatin Remodeling of Host Cell DNA via Epigenetic Mechanisms

Portions of this work have been published as a review to *Frontiers in Genetics* and is reprinted here with permission from Frontiers Media.

Sinclair SH, Rennoll-Bankert KE, and JS Dumler. 2014. Effector bottleneck: microbial reprogramming of parasitized host cell transcription by epigenetic remodeling of chromatin structure. *Front Genet*. 2014 Aug 14; 5:274

In order to infect mammalian hosts, bacterial pathogens have evolved an array of mechanisms that serve to create a hospitable environment conducive for survival, replication, and spread. While many bacterial species survive in an extracellular environment, intracellular pathogens must be capable of both entering their host cells undetected and altering the cellular milieu in order to replicate. Traditionally, the ability of bacterial-derived proteins to induce disruptions of cell signaling or major cellular processes such as NF-κB, MAPK, and JAK/STAT pathways, is the predominant focus of host-pathogen interaction studies.[1] Recently, there has been a renewed interest in the ability of these intracellular pathogens to induce alterations in host cell gene expression in order to promote survival and replication.[2-4] It is now well recognized that bacterial pathogens can reprogram host gene expression either directly or indirectly, by altering the accessibility of gene promoters via epigenetic modifications.

Eukaryotic DNA is highly organized and gene expression tightly regulated by an orchestrated network of proteins, RNAs and other modulators. Histone octamers, or nucleosomes, organize DNA by acting as a bobbin on which the DNA winds. The charge of the histone proteins can be covalently modified in order to more tightly or loosely associate with DNA. Histone acetylation, which imparts a negative charge, is predominantly associated with an open configuration where promoters are easily accessed by RNA polymerases. Histone methylation or phosphorylation, which impart a positive charge, cause DNA to more tightly associate with histone proteins, reducing promoter accessibility to transcription activating machinery. The process of modulating open and closed chromatin is further complicated by i) the residue(s) of which histone protein(s) is/are modified, ii) methylation of cytosine residues in DNA, and iii) non-

coding RNAs, and by the manner in which these mechanisms are intertwined. It therefore is no surprise that bacterial-derived proteins have evolved to interfere with host gene expression and improve bacterial fitness.

Over the last decade, examples of secreted bacterial effector proteins, ranging from those produced by *Listeria monocytogenes, Chlamydia trachomatis, Shigella flexneri,* and others, were found to target the host cell nucleus. These discoveries demonstrate the relationship of altered transcription and function of infected host cells and microbial survival, and in some cases, pathogenicity. *S. flexneri*, a bacterial pathogen that can cause dysentery, prevents NF-κB from binding its target gene promoters by altering the phosphorylation state of histone H3 at serine 10. The bacterium does so by secreting outer membrane protein F (OspF) that de-phosphorylates MAPKs in the nucleus resulting in a lack of histone H3 phosphorylation at serine 10 at a number of NF-κB-dependent genes.[5] By altering NF-κB target gene transcription, *S. flexneri* suppresses the host cell inflammatory response promoting microbial survival and transmission.[5] *L. monocytogenes* nuclear targeted protein A (LntA), blocks binding of heterochromatin inducing protein BAHD1 at interferon-stimulated genes resulting in upregulated expression.[6] Nuclear effector E (NUE) of *C. trachomatis*, and RomA of *L. pneumophila* have methyltransferase activity and induce methylation of eukaryotic histones and altered host cell gene expression.[7,8] These examples are part of an expanding array of bacterial-derived proteins termed nucleomodulins that target host cell chromatin or chromatin-linked pathways to alter transcription, typically at one or a few host genes. To date, the only prokaryotic nucleomodulins shown to directly bind DNA and influence surrounding chromatin are from the *Anaplasmataceae* family. Ankyrin A (AnkA) of *Anaplasma*

phagocytophilum, as well as Ank200 and several tandem-repeat containing proteins (TRPs) from *Ehrlichia chaffeensis*, have been shown to enter the nucleus and bind DNA, and interact with host epigenetic machinery or alter nearby histone octamers.[9-13]

Targeting of individual genes or binding of an effector at a single chromatin region or small numbers of promoter loci could, in theory, lead to c*is* regulation of those loci and in part explain transcriptional alterations induced by infection. However, the degree of transcriptional alterations and the coordination of these events that dramatically affect cellular function programs are unlikely to be explained by individual targets given the complexity and repertoire of the human genome, at approximately 3,251 Mb vs. genome sizes of *S. flexneri* (≤ 4.83 Mb),[14] *C. trachomatis* (≤ 1.04 Mb),[15] *L. monocytogenes* (≤3.00 Mb),[16] *L. pneumophila* (≤ 3.40 Mb),[17] and *A. phagocytophilum* (≤ 1.47 Mb).[18] We hypothesize that nucleomodulins can act broadly, even genome-wide, to affect entire chromosomal neighborhoods and topologically-associating chromatin domains by recruiting chromatin remodeling complexes or by altering the folding patterns of chromatin that bring distant regulatory regions together and reprogram and coordinate control of transcriptional programs. This review will discuss the current knowledge of *A. phagocytophilum* subversion of host cells by nucleomodulins and how AnkA could play an even larger role than immediately suspected.

References:

1. Brodsky IE, Medzhitov R. Targeting of immune signalling networks by bacterial pathogens. *Nat Cell Biol*. 2009;11 (5): 521-526.

2. Silmon de Monerri NC, Kim K. Pathogens hijack the epigenome: A new twist on host-pathogen interactions. *Am J Pathol*. 2014;184 (4): 897-911.

3. Bierne H, Hamon M, Cossart P. Epigenetics and bacterial infections. *Cold Spring Harb Perspect Med*. 2012;2 (12): a010272.

4. Paschos K, Allday MJ. Epigenetic reprogramming of host genes in viral and microbial pathogenesis. *Trends Microbiol*. 2010;18 (10): 439-447.

5. Arbibe L, Kim DW, Batsche E, et al. An injected bacterial effector targets chromatin access for transcription factor NF-kappaB to alter transcription of host genes involved in immune responses. *Nat Immunol*. 2007;8 (1): 47-56.

6. Rohde JR. Microbiology. *Listeria* unwinds host DNA. *Science*. 2011;331 (6022): 1271-1272.

7. Rolando M, Sanulli S, Rusniok C, et al. *Legionella pneumophila* effector RomA uniquely modifies host chromatin to repress gene expression and promote intracellular bacterial replication. *Cell Host Microbe*. 2013;13 (4): 395-405.

8. Pennini ME, Perrinet S, Dautry-Varsat A, Subtil A. Histone methylation by NUE, a novel nuclear effector of the intracellular pathogen *Chlamydia trachomatis*. *PLoS Pathog*. 2010;6 (7): e1000995.

9. Luo T, McBride JW. *Ehrlichia chaffeensis* TRP32 interacts with host cell targets that influence intracellular survival. *Infect Immun*. 2012;80 (7): 2297-2306.

10. Zhu B, Kuriakose JA, Luo T, et al. *Ehrlichia chaffeensis* TRP120 binds a G+C-rich motif in host cell DNA and exhibits eukaryotic transcriptional activator function. *Infect Immun*. 2011;79 (11): 4370-4381.

11. Luo T, Kuriakose JA, Zhu B, Wakeel A, McBride JW. *Ehrlichia chaffeensis* TRP120 interacts with a diverse array of eukaryotic proteins involved in transcription, signaling, and cytoskeleton organization. *Infect Immun*. 2011;79 (11): 4382-4391.

12. Dunphy PS, Luo T, McBride JW. *Ehrlichia* moonlighting effectors and interkingdom interactions with the mononuclear phagocyte. *Microbes Infect*. 2013;15 (14-15): 1005-1016.

13. Garcia-Garcia JC, Barat NC, Trembley SJ, Dumler JS. Epigenetic silencing of host cell defense genes enhances intracellular survival of the rickettsial pathogen *Anaplasma phagocytophilum*. *PLoS Pathog*. 2009;5 (6): e1000488.

14. Jin Q, Yuan Z, Xu J, et al. Genome sequence of *Shigella flexneri* 2a: Insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res*. 2002;30 (20): 4432-4441.

15. Stephens RS, Kalman S, Lammel C, et al. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science*. 1998;282 (5389): 754-759.

16. Evans MR, Swaminathan B, Graves LM, et al. Genetic markers unique to *Listeria monocytogenes* serotype 4b differentiate epidemic clone II (hot dog outbreak strains) from other lineages. *Appl Environ Microbiol*. 2004;70 (4): 2383-2390.

17. Chien M, Morozova I, Shi S, et al. The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science*. 2004;305 (5692): 1966-1968.

18. Lin M, Kikuchi T, Brewer HM, Norbeck AD, Rikihisa Y. Global proteomic analysis of two tick-borne emerging zoonotic agents: *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. *Front Microbiol*. 2011; 2:24.

Chapter 2: *Anaplasma phagocytophilum:* A model intracellular pathogen

Portions of this work have been published as a review to *Frontiers in Genetics* and is reprinted here with permission from Frontiers Media.

Sinclair SH, Rennoll-Bankert KE, and JS Dumler. 2014. Effector bottleneck: microbial reprogramming of parasitized host cell transcription by epigenetic remodeling of chromatin structure. *Front Genet*. Aug 14; 5:274.

Anaplasma phagocytophilum

Anaplasma phagocytophilum, transmitted by *Ixodes* spp. ticks, was discovered as the causative agent of human granulocytic anaplasmosis (HGA) in 1990.[1] While the infection is usually subclinical, manifestations in humans range from mild fever to severe infection requiring intensive care or even death.[1] *A. phagocytophilum* is a Gram negative, obligate intracellular bacterium that is a parasite of neutrophils in mammalian hosts, and infects a range of cells in the ticks that acquire it from and transmit it to mammals. Owing to their roles in protective inflammatory and immune responses toward microbial infections and their innate ability to recognize and kill pathogens, neutrophils are unlikely host cells for any microorganism. Yet, *A. phagocytophilum* colonizes these cells and thwarts their normal functions to create a hospitable environment for intracellular replication and subsequent transmission via tick bite. With a limited genome of approximately 1.5 Mb the potential genetic reservoir for controlling an infected host cell, whether in the tick or mammal, is only a fraction of the tick or human genome.[2] This simple observation suggests that to circumvent such extremes, its control mechanisms must be highly efficient, multifunctional, or target master regulators or similar checkpoints in eukaryotic cells.

One remarkable feature of *A. phagocytophilum*-infected cells is the marked change in transcriptional profiles that belies aberrant regulation of many key host pathways.[3-5] *A. phagocytophilum*-infected neutrophils are characterized by an "activated-deactivated" phenotype with major functional aberrations including decreased respiratory burst, delayed apoptosis, reduced transmigration across endothelial cell barriers, decreased antimicrobial activities such as phagocytosis and microbial killing, while

simultaneous increases in degranulation of vesicle contents including proteases and increased production of chemokines, are observed.[6-12] The net result of these changes is an increase in i) inflammatory responses that recruit new host cells, ii) prolonged survival of infected cells, iii) an inability to kill internalized microbes, and iv) net sequestration of infected cells within the intravascular compartment that is more readily accessed by the next tick bite.[13] In fact, interference with any of these processes using *in vitro* and *in vivo* models leads to reduced fitness of the microbe underscoring how important these functional changes are in mammalian hosts.

The biological basis for such dramatic changes in neutrophil function is increasingly studied by methods that range from examination of individual pathways to genome-wide system biology approaches. Transcriptional profiling of infected neutrophils and HL-60 cells, the latter a commonly used cell model for *A. phagocytophilum* infection, reveals altered transcription genome-wide, confirming that changes in transcription are not restricted to a few genes and limited cellular functions, but likely play a role in most of the known functional changes induced by infection.[4,5] Transcriptional downregulation of two components of the NADPH oxidase, *CYBB* encoding NOX2 or gp91^{phox} and *RAC2* encoding a GTPase, that assemble in the membranes of activated cells to generate superoxide production and promote microbial killing, plays a role in prolonged inhibition of respiratory burst.[6,14] In addition, delayed apoptosis is achieved by maintained transcription of *BCL2* family members[11] and increased proinflammatory responses are due to upregulated transcription of cytokine and chemokine genes such as *IL8.*[15]

While many studies document these changes at functional and transcriptional

levels, the precise mechanisms that organize and coordinate alterations to support improved *A. phagocytophilum* fitness are not well addressed. The biggest advance in understanding these processes occurred with the discovery of secreted effector proteins and their secretion systems. The genome of *A. phagocytophilum* encodes a type IV secretion system (T4SS) that is proven to allow bacterial protein effectors to translocate into infected host cells where they likely act by mechanisms similar to those of other bacterial secreted effectors that target cytosolic pathways such as signal transduction, cytoskeletal rearrangements, intracellular trafficking, etc.[16,17] These observations also hold true for *A. phagocytophilum* effectors. However, not all effector proteins remain localized within the host cytosol, and those that enter the nucleus have access to a distinct and diverse array of proteins that could impact cellular function on a global scale.

AnkA of *Anaplasma phagocytophilum*

A. phagocytophilum expresses a number of major immunoreactive proteins, including major surface protein 2/p44 (Msp2/p44) and Ankyrin A (AnkA), the latter of which was shown by Caturegli *et al.* using immunoelectron microscopy to be transported into the host cell nucleus and bound within heterochromatin.[18,19] AnkA contains many EPIYA motifs that become tyrosine phosphorylated and recruit SHP-1 and ABL1 when introduced into mammalian cells which in turn regulate endosomal entry and intracellular infection.[2,16] Aside from this, AnkA contains multiple eukaryotic motifs including 8-15 or more ankyrin repeats, a putative bipartite nuclear localization signal (NLS), and a putative high mobility group N-chromatin unfolding domain (HMGN-CHUD).[18] The ankyrin repeat domains are organized tandemly, creating highly stable spring-like structures that allow protein-protein or protein-DNA interactions and are commonly

found in transcription factors and their regulatory proteins.[20] HMGN-CHUD domains facilitate binding to nucleosomes to alter chromatin structure and transcription of surrounding genes.[21] The presence of these motifs lends credence to the idea that AnkA plays a role in altering neutrophil transcription.

In 2003, Park *et al* confirmed that AnkA directly binds both DNA and nuclear proteins, and provided limited evidence of its capacity to bind broadly throughout the human genome.[22] AnkA binds to regions at the promoter of *CYBB*, encoding the gp91^{phox} component of phagocyte oxidase.[10] *CYBB* is known to be transcriptionally repressed with *A. phagocytophilum* infection, further suggesting that AnkA acts in *cis* to alter gene transcription.[10,23] Moreover, transcription of *CYBB* is decreased in a dose dependent manner as nuclear AnkA concentrations increase.[10] When AnkA-expressing plasmids are transfected into HL-60 cells, *CYBB* expression is dampened, strengthening the link to transcriptional regulation.[10] In electrophoretic mobility shift assays, AnkA binds to the *CYBB* promoter at the same locations as other known transcriptional regulators such as CAATT displacement protein (CDP) and special-AT rich binding protein-1 (SATB1).[10,13] Surprisingly, DNA binding by AnkA does not target a conserved DNA sequence motif or signature; rather it binds to regions rich in AT nucleotides that have specific structural qualities, including the ability to uncoil under superhelical stress.[10,22] The latter feature is often observed in eukaryotic proteins that tether DNA strands to the nuclear matrix at matrix attachment regions (MARs) to regulate transcription from distantly-located but functionally-related chromosomal regions; proteins that bind these regions are called MAR-binding proteins (MARBPs).

Interestingly, Garcia-Garcia *et al.* showed that multiple downregulated defense

genes in infected granulocytes were clustered on chromosomes.[24] The spatial clustering of similarly regulated genes suggests that, in addition to *cis*-regulation as with *CYBB*, long ranges of chromosomes are affected – a process often attributed to epigenetic mechanisms such as DNA methylation and histone chromatin modifications. Using chromatin immunoprecipitation (ChIP) to investigate histone marks at defense gene promoters, acetylation of histone H3 was dramatically reduced, a finding often associated with silenced gene transcription.[24] To explain this, binding of histone deacetylase-1 (HDAC1) was found increased across multiple defense gene promoters.[24] Moreover, *A. phagocytophilum*-infected granulocytes have increased HDAC activity most likely explained by the increased quantity of both HDAC1 and HDAC2. Inhibition of *HDAC1*, but not *HDAC2* expression by siRNA or pharmacologic inhibition of HDAC activity impairs *A. phagocytophilum* propagation, whereas overexpression leads to increased intracellular propagation.[24] These results clearly link AnkA with HDAC1 toward facilitating widespread downregulation of antimicrobial responses. We modeled this process by using the wild type *CYBB* promoter, to which AnkA binds, and mutated forms unable to bind AnkA in order to demonstrate that AnkA binding leads to HDAC1 recruitment and silencing of expression at *CYBB*.[10]

References:

1. Chen SM, Dumler JS, Bakken JS, Walker DH. Identification of a granulocytotropic Ehrlichia species as the etiologic agent of human disease. *J Clin Microbiol*. 1994;32(3):589-595.

2. Lin M, Kikuchi T, Brewer HM, Norbeck AD, Rikihisa Y. Global proteomic analysis of two tick-borne emerging zoonotic agents: Anaplasma phagocytophilum and *Ehrlichia chaffeensis*. *Front Microbiol*. 2011;2:24.

3. de la Fuente J, Ayoubi P, Blouin EF, Almazan C, Naranjo V, Kocan KM. Gene expression profiling of human promyelocytic cells in response to infection with *Anaplasma phagocytophilum*. *Cell Microbiol*. 2005;7(4):549-559.

4. Borjesson DL, Kobayashi SD, Whitney AR, Voyich JM, Argue CM, Deleo FR. Insights into pathogen immune evasion mechanisms: *Anaplasma phagocytophilum* fails to induce an apoptosis differentiation program in human neutrophils. *J Immunol*. 2005;174(10):6364-6372.

5. Lee HC, Kioi M, Han J, Puri RK, Goodman JL. *Anaplasma phagocytophilum*-induced gene expression in both human neutrophils and HL-60 cells. *Genomics*. 2008;92(3):144-151.

6. Carlyon JA, Chan WT, Galan J, Roos D, Fikrig E. Repression of *rac2* mRNA expression by *Anaplasma phagocytophila* is essential to the inhibition of superoxide production and bacterial proliferation. *J Immunol*. 2002;169(12):7009-7018.

7. Carlyon JA, Fikrig E. Mechanisms of evasion of neutrophil killing by *Anaplasma phagocytophilum*. *Curr Opin Hematol*. 2006;13(1):28-33.

8. Carlyon JA, Fikrig E. Invasion and survival strategies of *Anaplasma phagocytophilum*. *Cell Microbiol*. 2003;5(11):743-754.

9. Dumler JS, Choi KS, Garcia-Garcia JC, et al. Human granulocytic anaplasmosis and *Anaplasma phagocytophilum*. *Emerg Infect Dis*. 2005;11(12):1828-1834.

10. Garcia-Garcia JC, Rennoll-Bankert KE, Pelly S, Milstone AM, Dumler JS. Silencing of host cell *CYBB* gene expression by the nuclear effector AnkA of the intracellular pathogen *Anaplasma phagocytophilum*. *Infect Immun*. 2009;77(6):2385-2391.

11. Ge Y, Yoshiie K, Kuribayashi F, Lin M, Rikihisa Y. *Anaplasma phagocytophilum* inhibits human neutrophil apoptosis via upregulation of *bfl-1*, maintenance of mitochondrial membrane potential and prevention of caspase 3 activation. *Cell Microbiol*. 2005;7(1):29-38.

12. Choi KS, Park JT, Dumler JS. *Anaplasma phagocytophilum* delay of neutrophil apoptosis through the p38 mitogen-activated protein kinase signal pathway. *Infect Immun*. 2005;73(12):8209-8218.

13. Rennoll-Bankert KE, Dumler JS. Lessons from *Anaplasma phagocytophilum*: Chromatin remodeling by bacterial effectors. *Infect Disord Drug Targets*. 2012;12(5):380-387.

14. Banerjee R, Anguita J, Roos D, Fikrig E. Cutting edge: Infection by the agent of human granulocytic ehrlichiosis prevents the respiratory burst by down-regulating *gp91phox*. *J Immunol*. 2000;164(8):3946-3949.

15. Klein MB, Hu S, Chao CC, Goodman JL. The agent of human granulocytic ehrlichiosis induces the production of myelosuppressing chemokines without induction of proinflammatory cytokines. *J Infect Dis*. 2000;182(1):200-205.

16. IJdo JW, Carlson AC, Kennedy EL. *Anaplasma phagocytophilum* AnkA is tyrosine-phosphorylated at EPIYA motifs and recruits SHP-1 during early infection. *Cell Microbiol*. 2007;9(5):1284-1296.

17. Ohashi N, Zhi N, Lin Q, Rikihisa Y. Characterization and transcriptional analysis of gene clusters for a type IV secretion machinery in human granulocytic and monocytic ehrlichiosis agents. *Infect Immun*. 2002;70(4):2128-2138.

18. Caturegli P, Asanovich KM, Walls JJ, et al. AnkA: An *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infect Immun*. 2000;68(9):5277-5283.

19. Kim HY, Mott J, Zhi N, Tajima T, Rikihisa Y. Cytokine gene expression by peripheral blood leukocytes in horses experimentally infected with *Anaplasma phagocytophila*. *Clin Diagn Lab Immunol*. 2002;9(5):1079-1084.

20. Bork P. Hundreds of ankyrin-like repeats in functionally diverse proteins: Mobile modules that cross phyla horizontally? *Proteins*. 1993;17(4):363-374.

21. Bustin M. Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol Cell Biol*. 1999;19(8):5237-5246.

22. Park J, Kim KJ, Choi KS, Grab DJ, Dumler JS. *Anaplasma phagocytophilum* AnkA binds to granulocyte DNA and nuclear proteins. *Cell Microbiol*. 2004;6(8):743-751.

23. Thomas V, Samanta S, Wu C, Berliner N, Fikrig E. *Anaplasma phagocytophilum* modulates *gp91phox* gene expression through altered interferon regulatory factor 1 and PU.1 levels and binding of CCAAT displacement protein. *Infect Immun*. 2005;73(1):208-218.

24. Garcia-Garcia JC, Barat NC, Trembley SJ, Dumler JS. Epigenetic silencing of host cell defense genes enhances intracellular survival of the rickettsial pathogen *Anaplasma phagocytophilum*. *PLoS Pathog*. 2009;5(6):e1000488.

Chapter 3: *Anaplasma phagocytophilum* induces genome-wide transcriptional changes in its host cell via chromatin remodeling

Portions of this work have been published as a review to *Frontiers in Genetics* and is reprinted here with permission from Frontiers Media.

Sinclair SH, Rennoll-Bankert KE, and JS Dumler. 2014. Effector bottleneck: microbial reprogramming of parasitized host cell transcription by epigenetic remodeling of chromatin structure. *Front Genet.* Aug 14; 5:274.

More than a *cis* interation

Currently, most data that examine the ability of HDAC1 and AnkA to modulate transcription focus on a small number of loci, including *CYBB* and up to 17 other gene promoters. While AnkA and HDAC1 are important contributors to transcriptional downregulation of some genes in *A. phagocytophilum*-infected cells, neither has been determined to provide a broad mechanism for transcriptional or functional alterations observed. Importantly, transcriptional profiling of *A. phagocytophilum*-infected granulocytes, including both primary neutrophils and cell lines, demonstrates that the majority of differentially expressed genes (DEGs) are upregulated.[1] Thus, while HDAC1 recruitment by AnkA is important for down-regulating some genes, the majority of DEGs are likely regulated by alternative or additional mechanisms.

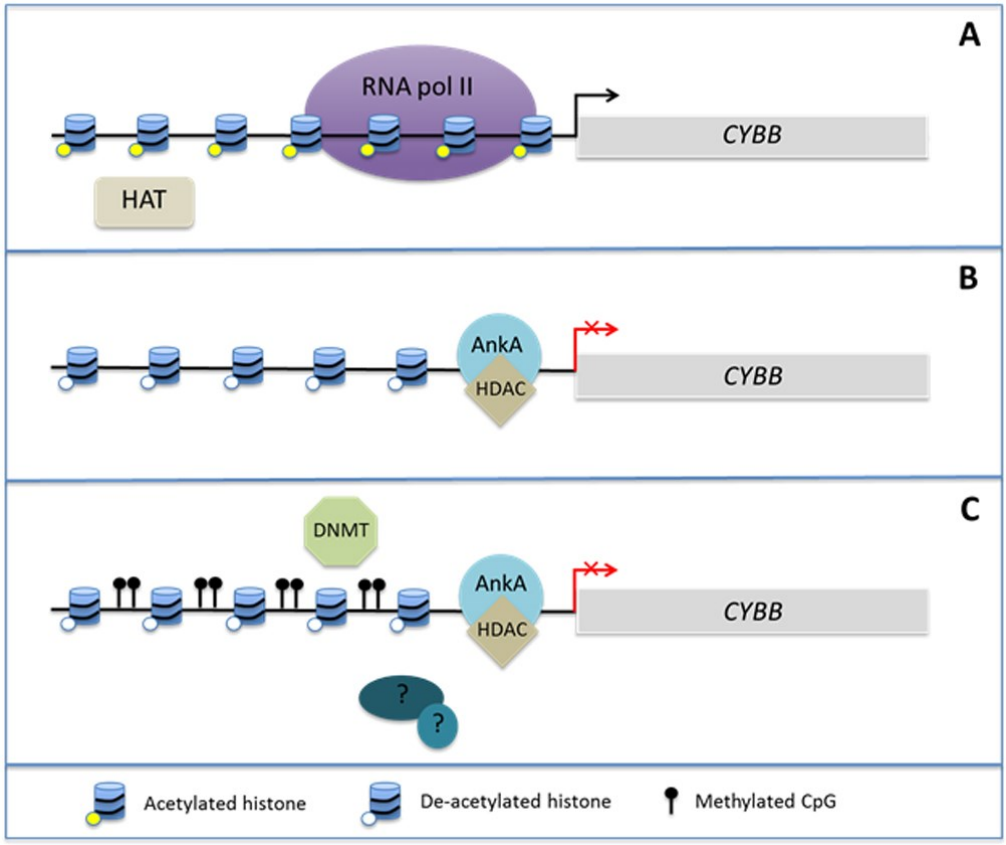
Of considerable interest is the interplay between chromatin structure and DNA methylation. Histone modifications, DNA methylation at CpG islands, and binding of methyl CpG DNA-binding proteins (MECPs) via their methyl-DNA binding domains (MBDs) occur in concert during biological responses including neoplasia and cellular differentiation.[2-6] Cytosine residues of CpG dinucleotides can by methylated by DNA methyltransferases (DNMTs) leading to sustained DNA methylation over multiple generations of cell divisions. It is believed in part that the methylation protects genes and transcriptional programs from inappropriate ectopic expression as cells enter various stages of tissue differentiation[5]. DNMT1 is highly conserved and is thought to be primarily responsible for maintaining existing methyl CpGs.[4] In contrast, DNMT3a and 3b are believed to mediate *de novo* methylation, induces CpG methylation in response to cellular signals and stimuli.[7] MBDs can recognize these newly methylated regions of

DNA and are often associated with HDACs in MeCP1 and MeCP2 complexes, illustrating the potential for a direct link between DNA methylation and alterations in chromatin structure.[8-10] Generally, hypermethylation of gene promoters occurs synchronously with histone deacetylation and decreased gene expression of many tumor suppressor genes.[2-4] This complex interplay has been best studied in cancer genomes, where it is currently unclear whether DNA methylation induces alterations in chromatin structure or if chromatin structure induces changes in DNA methylation. However, given the interplay between HDAC activity and DNA methylation, it is not unreasonable to hypothesize that *A. phagocytophilum* infection is associated with DNA methylation alterations of the host genome during the process of global transcriptome reprogramming (Figure 3.1).

In addition to chromatin alterations induced by HDACs, MARs are responsible for dictating the three-dimensional architecture of chromatin loops and serve as tethering points for DNA to the nuclear matrix.[11-13] Arranging chromatin into loops allows transcription factors attached to the matrix access to promoters and brings distal genomic loci and regulatory regions into a position of proximity for coordinated regulation.[14] Interestingly, the AT-rich DNA docking sites of AnkA are similar to those of MARs. In fact, several transcription factors including SATB1, bind MARs to coordinately modulate transcription from large genomic regions.[13] Like AnkA, SATB1 occupies the *CYBB* promoter during myeloid differentiation where it represses the transcription of *CYBB*, is involved in maintaining expression of *BCL2* [15] which is transcriptionally sustained during *A. phagocytophilum* infection,[16] and is implicated in activation of cytokine expression in T cells.[17,18] SATB1 interactions with chromatin remodeling complexes include histone

deacetylases which are likely involved in some of these processes.[19] The complex looping of chromatin facilitated by MARs and MAR binding proteins allows for entire chromatin domains and territories to be remodeled despite relatively few binding sites.[20] For example, SATB1 binds to only 9 regions across the 200 kB T_H2 locus, yet it is a major regulator of T_H2 lymphocyte differentiation and function.[20] Complex protein interactions involving chromatin remodeling and histone modifications mediated by anchoring proteins like SATB1 make it plausible that DNA binding bacterial nucleomodulins such as AnkA could target broad transcriptional programs that belie functions of host cells. While only limited data currently exist, AnkA binds to at least 23 distinct sites on 12 separate chromosomes with *A. phagocytophilum* infection of the human HL-60 promyelocytic cell line,[21] detailed mapping of AnkA genomic binding sites is now in progress. We therefore think it is plausible that AnkA binds throughout the genome and exerts its effects, like SATB1, to both repress and activate transcription by tethering chromatin to the nuclear matrix and exposing promoters to chromatin remodeling complexes.

Figure 3.1



AnkA alters chromatin structure at the *CYBB* promoter. During *A. phagocytophilum* infection, AnkA accumulates in the host cell nucleus where it can directly bind DNA and influence the transcription of *CYBB*, which encodes the gp91^{phox} component of the NADPH oxidase. A) In the absence of infection, *CYBB* is activated by inflammatory signals and is easily transcribed. B) During infection, AnkA binds the proximal promoter and recruits HDAC which deacetylates nearby histones in order to inhibit transcription. C) We propose that additional host or bacterial-derived chromatin modifying enzymes, such as DNMT methylation of host CpGs, may also be involved in altering the host epigenome to the bacterium's advantage.

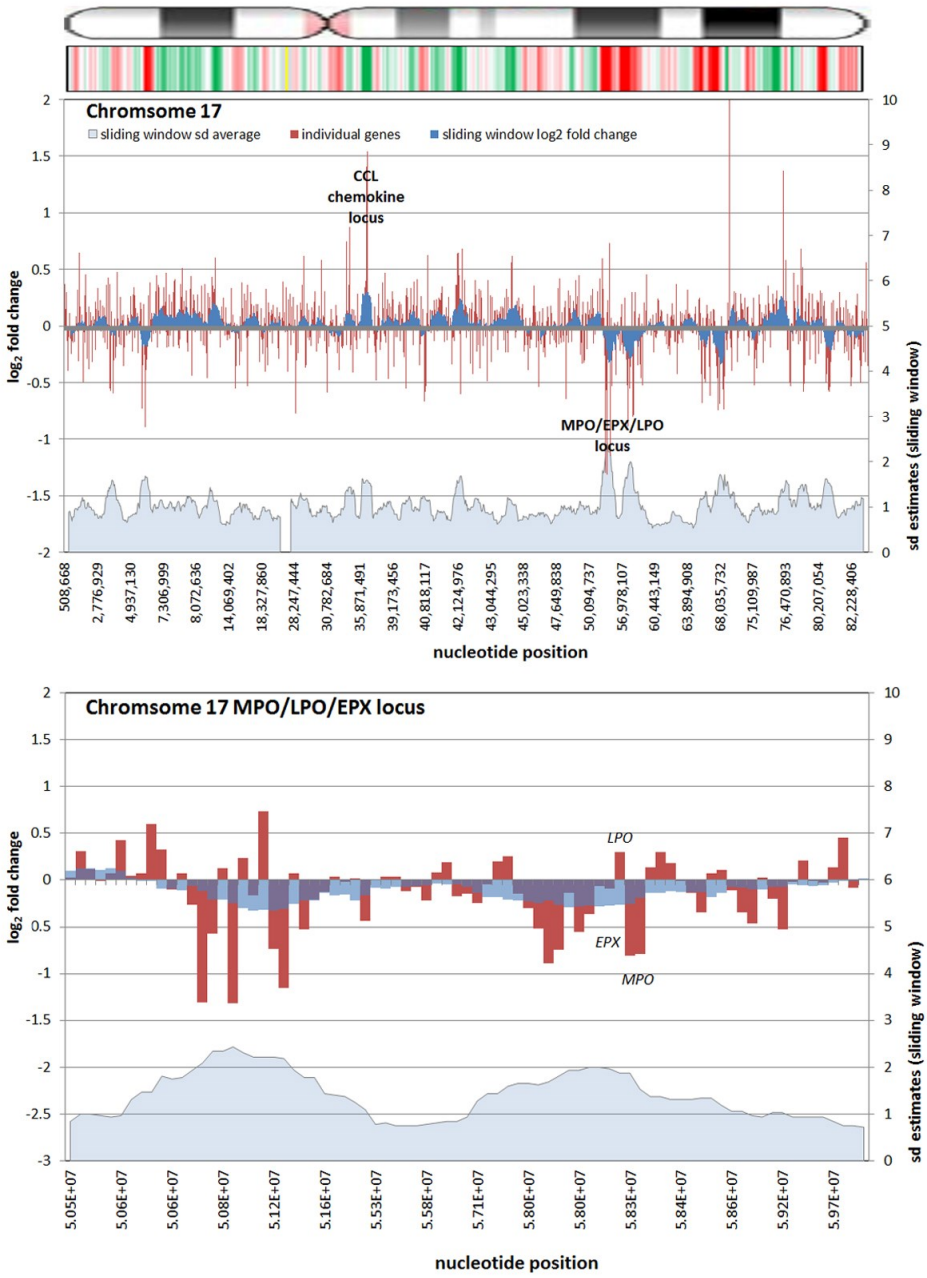
Anaplasma phagocytophilum infection is associated with transcriptional changes over megabases in the human genome

As a relatively new field, studies that examine the epigenome of cells infected by or as a consequence of bacterial infection, whether parasitic or symbiotic, are few and far between. Given the proven interaction of *A. phagocytophilum* AnkA with gene promoters, epigenetic alterations of nearby chromatin, and its MAR-binding attributes, we reanalyzed publicly available transcription microarray data generated from *A. phagocytophilum*-infected human peripheral blood neutrophils.[1] We used updated bioinformatics tools, including an analysis of the standard deviation estimates for differential transcription at each locus, and wherever possible, improved gene and gene feature localization using the NCBI GRCh38 human genome assembly (release date December 24, 2013). The differential expression of all 18,400 interrogated transcripts and variants, including 14,500 well-characterized human genes, was mapped by chromosomal position so that the relationship between the linear chromosome landscape and transcriptional activity could be visualized and investigated. To display physical locations on chromosomes with long regions of altered transcription, the fold change of expression for each locus was averaged with the nearest neighboring genes to create sliding windows for each chromosome. Overall, the median sliding window included 10.0 genes/gene features (IQR 2.0) spanning a median interval of 2.04 Mb (IQR 2.16) corresponding approximately to the lower end dimensions of chromosomal or gene expression dysregulation domains.[22] The sliding window average along with individual gene transcriptional fold changes were plotted for each chromosome. To assess significance of fold changes in the sliding window, a similar sliding window was created to

simultaneously show the average estimated standard deviations over each window.

All chromosomes showed linear regions of marked differential transcription covering megabases in scale on both p and q arms. In keeping with the original observations of Borjesson et al.,[1] most clusters were upregulated; however, the analysis also demonstrated a similar pattern of downregulated genes over long linear regions. Of interest was chromosome 17 with a 9.4 Mb cluster containing myeloid peroxidase (*MPO*), eosinophil peroxidase (*EPX*), and lactoperoxidase (*LPO*) which showed marked downregulation, and a 3.2 Mb cluster containing multiple upregulated chemokine genes (Figure 3.2). In addition to reduced transcript levels, histone 3 at *MPO*, and *EPX* promoters was previously shown to be deacetylated upon *A. phagocytophilum* infection.[23] Furthermore, a 3.6 Mb cluster on chromosome 6 containing *FLOT1, TNF*, and the major histocompatibility complex (MHC) loci was upregulated (Figure 3.3). The chromosomal landscape of the *MHC1* locus is well documented with respect to chromatin looping via matrix attachment regions (MARS) and histone marks.[24] It is difficult to determine with certainty whether these changes are the result of manipulation by microbial effectors like AnkA, or whether these represent host cellular responses to infection, or an amalgam of both. Regardless, these observations clearly suggest that *A. phagocytophilum* infection targets large chromosomal territories to induce transcriptional alterations, in addition to specific genes. Given that HL-60 cells transfected to express AnkA have nearly identical differential gene expression patterns (unpublished) and reductions in induced respiratory burst compared to *A. phagocytophilum*-infected cells, a hypothesis that prokaryotic nucleomodulins evolved to modulate cell function by epigenetic alterations is compelling.

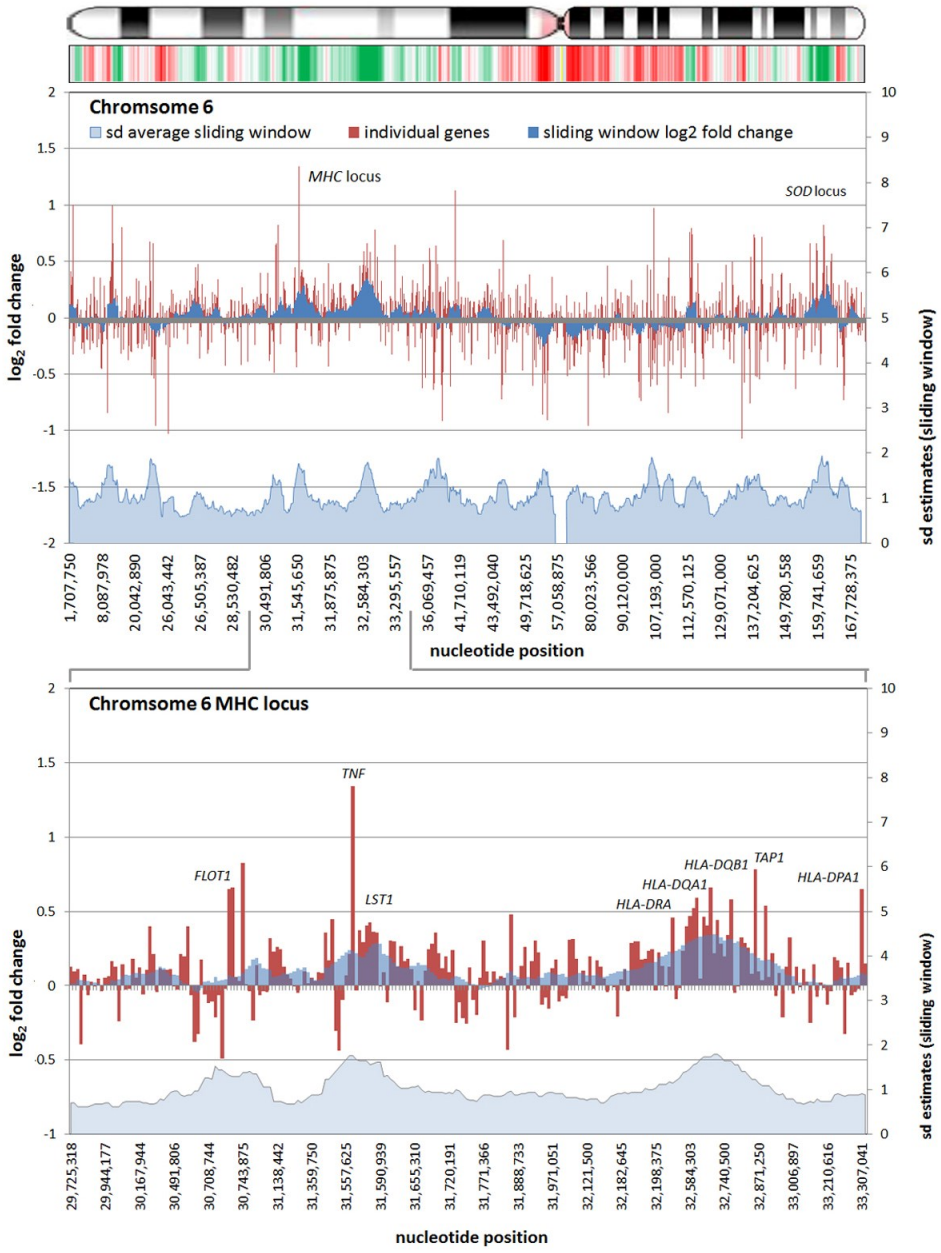
Figure 3.2



Differential gene expression patterns cluster over large linear genomic regions in *A. phagocytophilum*-infected human peripheral blood neutrophils. The top panel shows human chromosome 17; the bottom panel shows the *MPO*/*EPX*/*LPO* locus on human

chromosome 17 and the large genomic region that is upregulated with infection. Below the ideogram for the chromosome is a linear heat map for which dark green (upregulated) and dark red (downregulated) correspond to chromosomal positions where differential transcription exceeds 2 s.d. of the mean for chromosome 17. Red bars (left axes) represent differential transcription of individual genes, including some replicates; the dark blue zones (left axes) show the sliding window average $\log_2$-fold differential transcription over the contiguous 9-11 genes (or 0.38 to 1.25 Mb). The light blue zone at the bottom (right axis) shows the sliding window average over the same region for estimated standard deviations of $\log_2$-fold differential transcription at each gene or gene feature. Data re-analyzed from Borjesson et al. 2005[1] www.ncbi.nim.nih.gov/geo, accession no. GSE2405.

Figure 3.3



Differential gene expression patterns comparing *A. phagocytophilum*-infected vs. non-infected human peripheral blood neutrophils are clustered over large linear genomic regions. The top panel shows human chromosome 6; the bottom panel shows the MHC locus on human chromosome 6 including a large genomic region spanning the

TNF and *HLAD* loci that are upregulated with infection. Below the ideogram for the chromosome is a linear heat map for which dark green (upregulated) and dark red (downregulated) correspond to chromosomal positions where differential transcription exceeds 2 s.d. of the mean for chromosome 6. Red bars (left axes) represent differential transcription of individual genes, including some replicates, the dark blue zones (left axes) show the sliding window average $\log_2$-fold differential transcription over the contiguous 9-12 genes (or 0.42 to 3.24 Mb). The light blue zone at the bottom (right axis) shows the sliding window average over the same region for estimated standard deviations of $\log_2$-fold differential transcription at each gene or gene feature. Data re-analyzed from Borjesson et al. 2005[1] www.ncbi.nim.nih.gov/geo, accession no. GSE2405.

References:

1. Borjesson DL, Kobayashi SD, Whitney AR, Voyich JM, Argue CM, Deleo FR. Insights into pathogen immune evasion mechanisms: *Anaplasma phagocytophilum* fails to induce an apoptosis differentiation program in human neutrophils. *J Immunol*. 2005;174(10):6364-6372.

2. Stirzaker C, Song JZ, Davidson B, Clark SJ. Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. *Cancer Res*. 2004;64(11):3871-3877.

3. Baylin SB, Herman JG. DNA hypermethylation in tumorigenesis: Epigenetics joins genetics. *Trends Genet*. 2000;16(4):168-174.

4. Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum Mol Genet*. 2001;10(7):687-692.

5. Srivastava S, Mishra RK, Dhawan J. Regulation of cellular chromatin state: Insights from quiescence and differentiation. *Organogenesis*. 2010;6(1):37-47.

6. Reddington JP, Pennings S, Meehan RR. Non-canonical functions of the DNA methylome in gene regulation. *Biochem J*. 2013;451(1):13-23.

7. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. 1999;99(3):247-257.

8. Jones PL, Veenstra GJ, Wade PA, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet*. 1998;19(2):187-191.

9. Nan X, Ng HH, Johnson CA, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*. 1998;393(6683):386-389.

10. Ng HH, Zhang Y, Hendrich B, et al. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet*. 1999;23(1):58-61.

11. Spector DL. The dynamics of chromosome organization and gene regulation. *Annu Rev Biochem*. 2003;72:573-608.

12. Kohwi-Shigematsu T, Kohwi Y, Takahashi K, et al. SATB1-mediated functional packaging of chromatin into loops. *Methods*. 2012;58(3):243-254.

13. Kumar PP, Bischof O, Purbey PK, et al. Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat Cell Biol*. 2007;9(1):45-56.

14. Arope S, Harraghy N, Pjanic M, Mermod N. Molecular characterization of a human matrix attachment region epigenetic regulator. *PLoS One*. 2013;8(11):e79262.

15. Gong F, Sun L, Sun Y. A novel SATB1 binding site in the *BCL2* promoter region possesses transcriptional regulatory function. *J Biomed Res*. 2010;24(6):452-459.

16. Ge Y, Yoshiie K, Kuribayashi F, Lin M, Rikihisa Y. *Anaplasma phagocytophilum* inhibits human neutrophil apoptosis via upregulation of *bfl-1*, maintenance of mitochondrial membrane potential and prevention of caspase 3 activation. *Cell Microbiol*. 2005;7(1):29-38.

17. Hawkins SM, Kohwi-Shigematsu T, Skalnik DG. The matrix attachment region-binding protein SATB1 interacts with multiple elements within the *gp91phox* promoter and is down-regulated during myeloid differentiation. *J Biol Chem*. 2001;276(48):44472-44480.

18. Beyer M, Thabet Y, Muller RU, et al. Repression of the genome organizer SATB1 in regulatory T cells is required for suppressive function and inhibition of effector differentiation. *Nat Immunol*. 2011;12(9):898-907.

19. Yasui D, Miyano M, Cai S, Varga-Weisz P, Kohwi-Shigematsu T. SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature*. 2002;419(6907):641-645.

20. Cai S, Lee CC, Kohwi-Shigematsu T. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet*. 2006;38(11):1278-1288.

21. Park J, Kim KJ, Choi KS, Grab DJ, Dumler JS. *Anaplasma phagocytophilum* AnkA binds to granulocyte DNA and nuclear proteins. *Cell Microbiol*. 2004;6(8):743-751.

22. Letourneau A, Santoni FA, Bonilla X, et al. Domains of genome-wide gene expression dysregulation in down's syndrome. *Nature*. 2014;508(7496):345-350.

23. Garcia-Garcia JC, Barat NC, Trembley SJ, Dumler JS. Epigenetic silencing of host cell defense genes enhances intracellular survival of the rickettsial pathogen *Anaplasma phagocytophilum*. *PLoS Pathog*. 2009;5(6):e1000488.

24. Ottaviani D, Lever E, Mitter R, et al. Reconfiguration of genomic anchors upon transcriptional activation of the human major histocompatibility complex. *Genome Res*. 2008;18(11):1778-1786.

Chapter 4: Global DNA methylation changes and differential gene expression in *Anaplasma phagocytophilum*-infected neutrophils.

Manuscript submitted to *Genome Biology* in March 2015.

Sinclair SH, Yegnasubramanian S, Dumler JS. 2015. Global DNA methylation changes and differential gene expression in *Anaplasma phagocytophilum*-infected neutrophils.

Abstract

Anaplasma phagocytophilum is an obligate intracellular prokaryotic pathogen that both infects and replicates within human neutrophils. The bacterium represses multiple antimicrobial functions while simultaneously increasing proinflammatory functions by reprograming the neutrophil genome. Previous reports show that many observed phenotypic changes are in part explained by altered gene transcription. We recently identified that large chromosomal regions of the neutrophil genome are differentially expressed during *A. phagocytophilum* infection. Because of this, we sought to determine whether gene expression programs altered by infection were the result of changes in the host neutrophil DNA methylome. Within 24 hours of infection, marked increases in DNA methylation were observed genome-wide as compared with mock-infected controls and, pharmacologic inhibition of DNA methyltransferases resulted in decreased bacterial growth. New regions of DNA methylation were enriched at intron and exon junctions; however, intragenic methylation did not correlate with altered gene expression. In contrast, intergenic DNA methylation was associated with *A. phagocytophilum*-induced gene expression changes. Within the major histocompatibility complex locus on chromosome 6, a region with marked changes in infection-induced differential gene expression, new regions of methylation were localized to boundaries of active and inactive chromatin. These data strongly suggest that *A. phagocytophilum* infection, in addition to altering histone structure, alters DNA methylation and the epigenome of its host cell to promote survival and replication, providing evidence that such bacterial infection can radically alter the epigenome of its host cell.

Introduction

Anaplasma phagocytophilum is an obligate intracellular prokaryotic pathogen of mammalian neutrophils. Transmitted by *Ixodes* spp. ticks, *A. phagocytophilum* infection results in manifestations ranging from subclinical infection or mild self-limited fever to severe infection requiring hospitalization or causing death. [1] The neutrophil is an unusual host cell for any microorganism, especially bacteria, due to its role in innate immunity where its primary function is to recognize and kill microbes. Despite this inhospitable environment, *A. phagocytophilum* reprograms its host cell so that the bacterium can survive long enough for continued replication. The major phenotypic alterations among infected human neutrophils include complex functions such as decreased antimicrobial activity and respiratory burst, reduced phagocytosis, reduced margination and emigration across the endothelium, delayed apoptosis and an increased production and release of proinflammatory cytokines, chemokines, and proteases. [2-7] These functional alterations allow the bacterium to survive long enough to replicate and spread to newly recruited host cells, and to ensure that its host cell remains viable within the intravascular network until it can be accessed by a feeding tick.

The observed changes in host cell function are partially explained by alterations in granulocyte gene transcription. [8-10] Decreased respiratory burst allows for prolonged inhibition of superoxide production favoring bacterial survival and likely results from the downregulation of two key components of the NADPH oxidase required for proper oxidase assembly: *CYBB*, which encodes gp91^{phox}, and *RAC2*, a GTPase. [2,11] Likewise, transcription of *BCL2* family member genes are maintained during *A. phagocytophilum* infection which results in delayed neutrophil apoptosis, [6,12] and increased transcript levels

of cytokines and chemokines such as IL-1α and IL-8 contribute to exaggerated inflammatory responses that recruit new neutrophil hosts. [13-15]

Precisely how *A. phagocytophilum* coordinates the transcriptional changes that allow it to reprogram host cell functions is not known but is unlikely to result from of an accumulation of individual signaling pathways targeted by unique pathogen protein effectors. We described decreased histone acetylation and methylation at multiple defense gene promoters with infection, [16] and enhanced intracellular bacterial growth with HDAC-1 and -2 expression and activity, as well as the dose-dependent decrease in bacterial load effects of HDAC1 silencing and HDAC pharmacologic inhibition. [16] We also showed that transfection of the nucleomodulin and type IV system secretion substrate AnkA of *A. phagocytophilum*, alone mimics many transcriptional changes observed with infection by directly binding host cell DNA in multiple genomic regions, including at the *CYBB* proximal promoter where it decreases histone H3 deacetylation and gene expression. [17] Previously, we showed that AnkA recruits HDAC1 to the *CYBB* promoter, and the altered chromatin configuration excludes RNA polymerase II recruitment thereby silencing expression. [17] These observations prove an epigenetic basis for silencing at *CYBB* and suggest that *A. phagocytophilum* infection alters transcriptional programs of its host cells via genome-wide epigenetic alterations.

HDACs target nucleosomal histone proteins to create modifications that alter gene promoter accessibility to transcriptional regulators. Interestingly, HDACs also complex with DNA methyl binding proteins (MBDs) and work to coordinately alter transcriptional programs. DNA methylation at CpG dinucleotides is an important epigenetic regulator of cellular differentiation, neoplasia, and metastasis. [18-21] Moreover, expression of DNA

methyltransferase 3A (DNMT3A) a key enzyme for de novo DNA methylation, is upregulated in *A. phagocytophilum*-infected neutrophils *ex vivo*. [8] Given the complex interplay between these two epigenetic modifiers and the induction of histone deacetylation by *A. phagocytophilum*, we hypothesized that *A. phagocytophilum* also alters the DNA methylome of human neutrophils that in part determines transcriptional reprogramming.

Here, we found that *A. phagocytophilum* induces hypermethylation of the human neutrophil genome within 24 h of infection. This is the first demonstration that infection by an intracellular bacterium alters host DNA methylation patterns on a genome-wide scale. The data strongly suggest that this is part of a coordinated reprogramming of host cell functions which leads to improved microbial fitness by promoting intracellular survival, replication, and spread.

Methods

Cell Lines and Cell Culture

Primary peripheral blood neutrophils were isolated from venous blood of 3 healthy adult donors (2 females, 1 male) as approved by the Johns Hopkins Medicine IRB and as previously described. [22] Briefly, EDTA anticoagulated blood was dextran-sedimented and leukocyte-rich plasma centrifuged though a Ficoll-Paque gradient. Mononuclear cells were removed and the remaining erythrocytes were lysed in hypotonic saline. HL-60 promyelocytic cells (ATCC CCL-240) were purchased from American Type Tissue Culture (Manassas, VA). Both neutrophils and HL-60 cells were maintained or propagated in RMPI 1640 (Hyclone, Thermo Fisher Scientific, Waltham MA) with 10% fetal bovine serum (Thermo Fisher Scientific) and Glutamax (Life Technologies, Carlsbad CA). HL-60 cells were differentiated for 5 days prior to infection with 1 μM all-*trans* retinoic acid (ATRA). All cells were grown at 37°C in a humidified incubator with 5% CO_2. Where stated, HL-60 cells were treated with 5-azacytidine (Sigma-Aldrich, St. Louis MO) in DMSO for 24h or DMSO alone as a vehicle control.

Anaplasma phagocytophilum propagation and extraction

A. phagocytophilum was maintained in HL-60 cells as previously described. [23] Cell-free bacteria were obtained from HL-60 cells of which 90-95% were infected. Infected cells were centrifuged at 500 × g for 5 min and resuspended in 1× PBS. Infected cells were lysed by sonication using an output of 4 of a Branson Sonifier 250 (Branson Ultrasonics, Danbury CT) for 15 sec or via syringe lysis using a 22 g syringe. Lysed cells were then centrifuged at 1000 × g for 10 min and the bacteria-enriched supernatant was centrifuged at 13,000 × g for 30 min. The bacterial pellets were suspended in RMPI 1640

medium.

Neutrophils were infected *ex vivo* in triplicate using a multiplicity of infection (MOI) of 25:1 and 100:1 for experiments using HL-60 cells. An estimate of 10 infectious bacteria per infected HL-60 cell was used. The proportion of infected cells was determined at 24 h after Romanowsky staining (Protocol HEMA3, Thermo Fisher Scientific) of cytocentrifuged cells. In addition, DNA extracted from these cultures (DNeasy Blood and Tissue Kit, Qiagen, [Valencia CA]) was used in a quantitative real time 5' nuclease PCR assay that targets *A. phagocytophilum msp2/p44* normalized to human *ACTB*. [15] A standard curve of cloned *msp2* was used to accurately estimate the number of bacteria per cell, providing that the *A. phagocytophilum* genome encodes approximately 100 *msp2/p44* genes.

MBD-Seq library preparation

Enrichment of 5-methylcytosine modified DNA was performed using DNA extracted from infected and uninfected neutrophils 24 h post *ex vivo* infection. DNA libraries isolated by MBD enrichment were analyzed by MBD-seq to determine sites of MBD protein binding and DNA methylation (SKCCC Next Generation Sequencing Core). MBD-seq was carried out as described previously. [24] Briefly, 2 µg of DNA was sonicated to an average molecular size of ~150-250 bp, and end-repaired using the NEBNext SOLiD DNA library preparation kit end-repair module following the manufacturer's protocol (New England Biolabs, Ipswich MA). Fragments were purified using a Qiagen PCR purification kit (Qiagen). SOLiD P1 and P2 adapters lacking 5' phosphate groups (Life Technologies) were ligated to the fragment using the NEBNext

adapter ligation module, column-purified, and subjected to isothermal nick-translation by treating with Platinum Taq polymerase to remove the nick. The resulting library was divided into two fractions, a total input fraction, and an enriched methylated fraction. The enriched methylated fraction was subjected to affinity enrichment of methylated DNA fragments using recombinant C-terminal 6xHis-tagged MBD2-MBD polypeptides immobilized on magnetic beads, similar to previously described methods. [25-27] The resulting enriched methylated fraction and the total input fraction were then subjected to library amplification using the NEBNext amplification module according to the manufacturer's protocols, using 4 - 6 cycles for the total input, and 10 - 12 cycles for the enriched methylated fractions. Library fragments between 200 - 300 bp were selected after agarose gel electrophoresis. The libraries were then subjected to emulsion PCR and bead enrichment following the SOLiD emulsion PCR protocol (Life Technologies). The resulting beads were then deposited on the SOLiD flow cell and subjected to massively parallel 50 bp single-read sequencing on a Life Technologies SOLiD sequencer. Reads were mapped to the human genome hg18 build using Bioscope software. The total number of reads per sample, the % alignment are summarized in Supplemental Table 4.1. Regions of methylated DNA enrichment were identified using MACS [28] and SICE [29] as described previously. [30] Supplemental Table 4.2 summarizes the number of peaks called with both MACS and SICER as well as delineate the total number of peaks above a given threshold of -10*log10(p-value) >50, >100, and >200.

Supplemental Table 4.1

	# reads	% Aligned*
PMN.Donor 1_Enrich	26,653,200	74.34%
PMN.Donor 1_Total	5,654,374	92.53%
PMN.Donor 2_Enrich	16,167,853	82.18%
PMN.Donor 2_Total	6,510,552	90.66%
PMN.Donor 3_Enrich	24,553,732	81.39%
PMN.Donor 3_Total	6,450,849	87.42%
PMN.Aph.Donor 1_Enrich	154,514,937	82.03%
PMN.Aph.Donor 1_Total	19,913,892	87.36%
PMN.Aph.Donor 2_Enrich	63,786,363	84.18%
PMN.Aph.Donor 2_Total	4,708,289	92.02%
PMN.Aph.Donor 3_Enrich	25,186,600	81.49%
PMN.Aph.Donor 3_Total	15,855,721	91.20%

Genome coverage. Genome coverage as dictated by the total number of reads for enriched and total input fractions. The % alignment was calculated as the percentage of reads that aligned to the hg18 human reference genome.

Supplemental Table 4.2.

	MACS	>50	>100	>200
PMN.Donor 1	77,611	77,611	46,712	16,628
PMN.Donor 2	81,265	81,256	45,361	14,575
PMN.Donor 3	85,088	85,088	51,194	20,407
PMN.Aph.Donor 1	226,428	226,428	161,699	89,895
PMN.Aph.Donor 2	175,299	175,299	117,457	55,552
PMN.Aph.Donor 3	111,010	111,010	92,558	41,802

	SICER	>50	>100	>200
PMN.Donor 1	61,060	58,620	49,533	30,786
PMN.Donor 2	66,938	64,418	53,083	29,141
PMN.Donor 3	93,142	86,206	72,248	35,960
PMN.Aph.Donor 1	162,611	156,391	144,901	113,160
PMN.Aph.Donor 2	144,902	134,686	118,849	70,021
PMN.Aph.Donor 3	104,810	100,947	82,704	43,113

Number of peaks called using increasing stringency. Significant peaks were called using MACS v1.4 and SICERV1.1. Significant peaks were sorted by the -10*log10(p-value) score based on increasing stringencies to determine the number of peaks present above thresholds of >50, >100 and >200.

Luminometric Methylation Assay (LUMA)

LUMA was performed by EpigenDx (Hopkinton, MA) as described by Karimi *et al*.[31] Briefly, samples were digested with *Hpa*II + *Eco*RI or *Msp*I + *Eco*RI and fragments were amplified by Pyrosequencing™ (Qiagen). The ratio of (dGTP + dCTP)/dATP was calculated for *Hpa*II/*Eco*RI and *Msp*I/*Eco*RI fragments and the percentage of methylation determined by the ratio of (*Hpa*II/*Eco*RI)/(*Msp*I/*Eco*RI). Samples were tested in duplicate.

Identifying gene regions enriched for DNA methylation

Publically available *cis*-regulatory element annotation (CEAS) software [32,33] was used to characterize regions of methylated DNA enrichment by calculating the fraction present in different gene regions (introns, exons, 5' UTR, 3'UTR, and distal intergenic regions), promoters, bi-directional promoters, and regions downstream of gene bodies; this was also used to create average profile analyses of the smoothed adjusted $\log_2$ (M/T) values across transcriptional start sites (TSS), transcriptional termination sites (TTS), long (2,715-11,673 bp), medium (842-2,715 bp), and short (158-842 bp) intronic sequences, and long (164-483 bp), medium (109-164 bp), and short (66-109 bp) exonic sequences.

Identifying chromosomal locations enriched for DNA methylation and differential gene expression

Publically available transcription microarray data from *A. phagocytophilum*-

infected neutrophils was re-analyzed using RMA. [8] Only the 24 h time point was used for comparison with the same interval in the DNA methylation studies. Differential expression was determined based on the standard deviation of the fold change from the mean. Infected: uninfected fold change gene expression data were transformed to $\log_2$ for all analyses.

Regions enriched for DNA methylation and differential expression were organized linearly among their respective chromosomes and divided into bins of ~ 1 million base pairs. The number of gene features (expression) and methylation peaks within each bin were counted and compared to the average number across the genome to determine locations of interest. Regions of interest were defined by those bins which had significantly more counts than average for that chromosome.

DNA methylation and transcription in *A. phagocytophilum*-infected neutrophils

Because *A. phagocytophilum* infection induced genome-wide DNA hypermethylation, we investigated whether this affected gene transcription. Borjesson *et al* previously investigated the effect of *A. phagocytophilum* infection on transcription of the neutrophil using microarrays.[8] To access and compare these data to the DNA methylation analyses, the transcriptional profiling data were obtained from Gene Expression Omnibus (GEO accession number GSE2405) and re-normalized using RMA, as described above. Differential expression was determined for 24 h post infection, and genes were sorted by chromosome and physical location based on annotations in human genome assembly hg18 used for both these data and the DNA methylation studies.

Results:

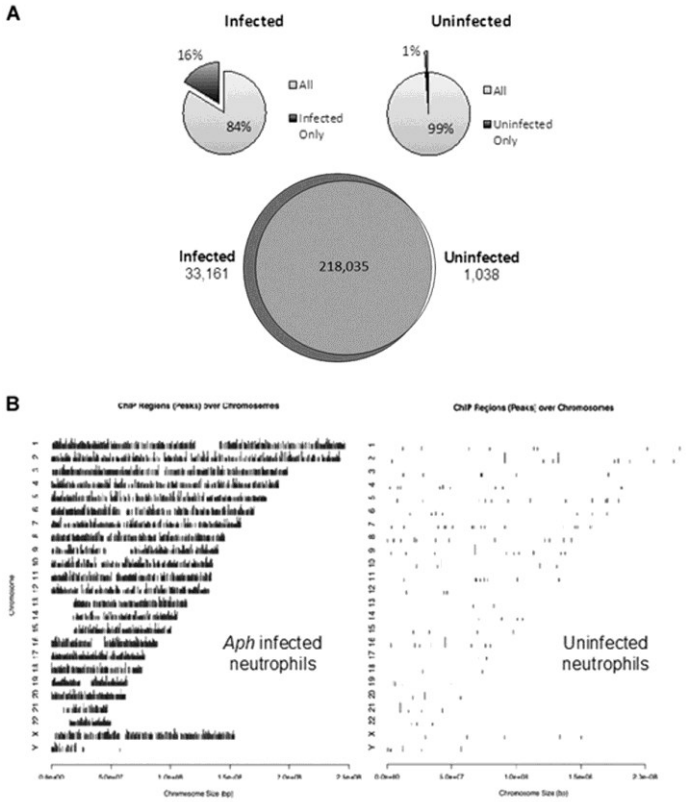
A. phagocytophilum induces hypermethylation genome-wide.

Human peripheral blood neutrophils were isolated and infected with *A. phagocytophilum* for 24 h before DNA isolation. While the majority of methylated regions (~86%) were maintained between uninfected and infected neutrophils, there were a significant number of regions (~30,000) of hypermethylation in the infected compared to uninfected samples (Figure 4.1A). Regions of methylation unique to infected or uninfected samples were mapped linearly along chromosomes. This showed that the changes in DNA methylation associated with infection were pervasive across all chromosomes (Figure 4.1B) and confirmed that alterations in DNA methylation patterns occurred on a genome-wide scale. LUMA analysis of the same DNA samples further confirmed an increase in the percentage of DNA methylation with *A. phagocytophilum* infection of human neutrophils.

Inhibition of DNMTs with 5-azacytidine slows *A. phagocytophilum* growth.

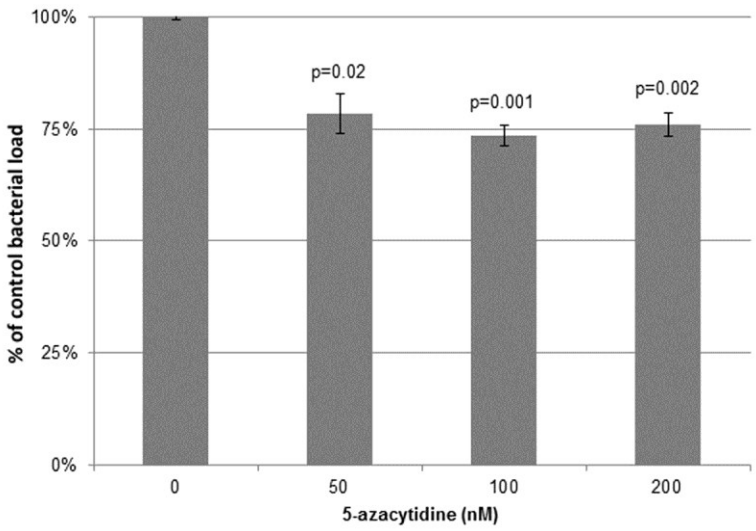
In concert with increased global DNA methylation, transcription of *DNMT3A* is upregulated with *A. phagocytophilum* infection of human neutrophils *ex vivo*.[8] To determine whether methylation of host cell DNA is essential for bacterial survival and growth, we investigated the pharmacologic effects of the DNA methyltransferase inhibitor 5-azyctidine on *A. phagocytophilum* infection. ATRA-differentiated HL-60 promyelocytic cells, commonly used as a cell model for *A. phagocytophilum* infection, were infected with cell-free *A. phagocytophilum* for 24 h before treatment with 5-azayctidine. After 24 h, bacterial replication was ~25% less than the DMSO vehicle control with all concentrations of 5-azacytidine (Figure 4.2).

Figure 4.1



Anaplasma phagocytophilum **induces genome-wide changes in DNA methylation of human peripheral blood neutrophils.** The quantity of new regions of DNA methylation were compared between infected and uninfected samples. A) Pie charts demonstrate the portion of new methylation peaks mapped to infected cells only versus uninfected cells. While a large percentage of peaks are shared between samples, infected samples have more new regions of methylation as shown by the Venn diagram. B) CEAS was used to map chromosomal locations of all methylated peaks unique to infected or uninfected samples. Representative image of a single donor where regions of DNA methylation (bars) unique to *A. phagocytophilum*-infected neutrophils and to uninfected human neutrophils are shown per chromosome on the y-axis, by their linear location on the x-axis.

Figure 4.2



DNA methyltransferase inhibition by 5-azacytidine abrogates *A. phagocytophilum* growth in HL-60 cells. Treatment for 24 h with the DNMT inhibitor 5-azacytidine 24 h post *A. phagocytophilum* infection results in a 25% reduction in bacterial load regardless of inhibitor concentration. Results are normalized to HL-60 cell *ACTB* and to vehicle treated cells. P values displayed above the bars were calculated by two-sided unequal variance Student's t-tests by comparison with untreated samples.

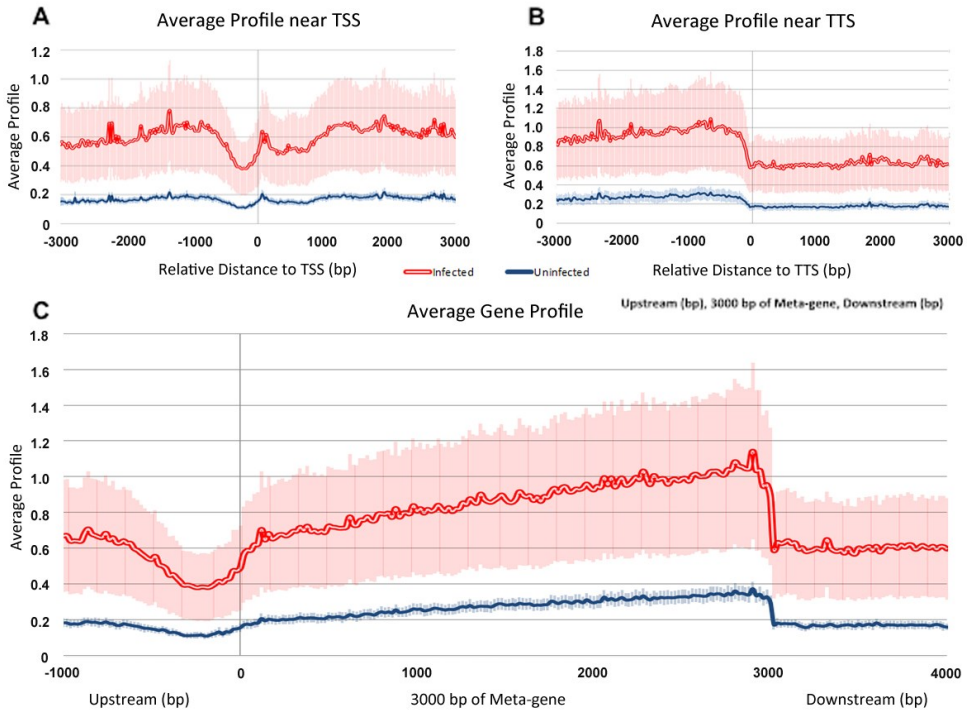
Characterization of DNA methylation across common gene features.

We next sought to determine which gene features were most prominently methylated throughout the genome with infection. The average methylation signal surrounding all transcription start sites (TSS), termination sites (TTS) and across genes for each individual donor and are shown in Figure 4.3. Both *A. phagocytophilum*-infected and uninfected samples showed similar relative profiles of DNA methylation across gene features. However, compared to uninfected, the profiles from *A. phagocytophilum*-infected samples showed a more pronounced decrease in DNA methylation immediately upstream of the TSS (Figure 4.3A), whereas downstream of the TSS, DNA methylation rose sharply and continued to increase throughout the gene body until the TTS (Figure 4.3B-C). These differences could not be accounted for by absolute numbers of sequencing reads.

Because infection with *A. phagocytophilum* induced greater spread and degrees of DNA methylation compared to uninfected cells, we investigated whether the newly methylated regions localized to specific gene features. Regions of methylation were highly associated with genes in *A. phagocytophilum*-infected neutrophils and uninfected neutrophils. Nearly 62 and 66%, respectively, mapped to or within 3 kb of annotated genes. Infected samples had higher than expected methylation enrichment within 2-3 kb upstream or downstream of genes (Figure 4.4) as well as increased methylation of exons (Figure 4.5). Nearly 5% of all newly methylated regions in *A. phagocytophilum*-infected cells fell within exons compared to the expected value of 2% (Figure 4.5). Expected values were determined by CEAS for what percentage of the human genome maps to indicated areas in the absence of enrichment.

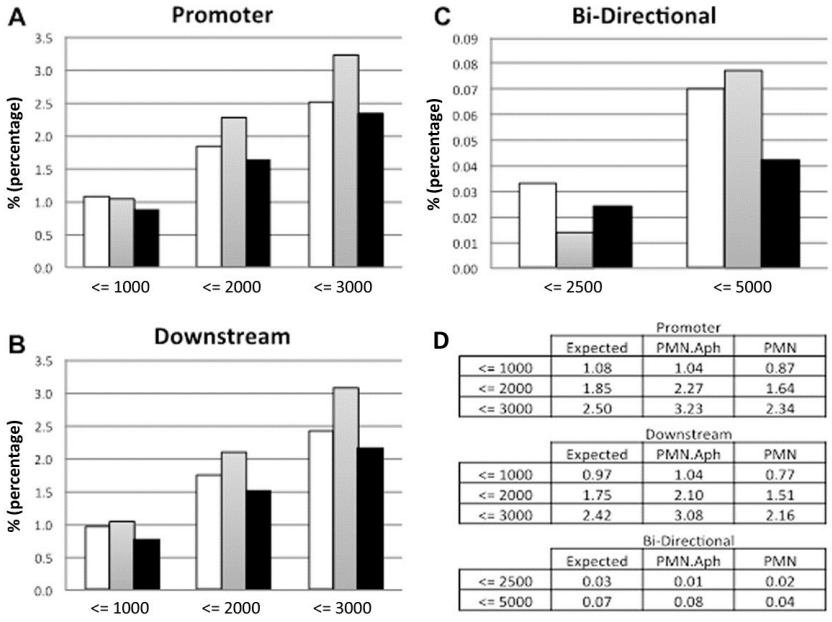
Intron and exon methylation was averaged, scaled as a fraction between 0 to 100%, and plotted by position. The profile for all sizes of introns calculated (small, medium, and large) had regions of methylation localizing toward the intron/exon junctions (Figure 4.6A). For all sizes of exons calculated (small, medium, and large) there was a gradual accumulation of methylation signals toward the middle of the exons (Figure 4.6B). The pattern of both profiles suggests that exons could be preferentially methylated and regions of methylation could overlap at intron/exon junctions. Overall, *A. phagocytophilum* infection-induced hypermethylation of the host genome leads to regions of enrichment most often associated with gene bodies, or within 3 kb of annotated gene features.

Figure 4.3



Average signal profile for 5mC enrichment. A) transcriptional start sites (TSS), B) transcription termination sites (TTS), and C) across an average gene was calculated using CEAS. Light gray represents the average profile of enrichment for *A. phagocytophilum* infected peripheral blood neutrophils, and dark gray for uninfected. Shaded zones show the standard error of the mean for each donor.

Figure 4.4



D

Promoter

	Expected	PMN.Aph	PMN
<= 1000	1.08	1.04	0.87
<= 2000	1.85	2.27	1.64
<= 3000	2.50	3.23	2.34

Downstream

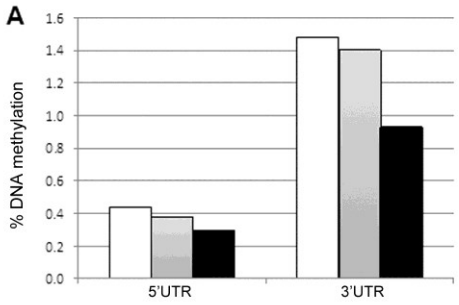
	Expected	PMN.Aph	PMN
<= 1000	0.97	1.04	0.77
<= 2000	1.75	2.10	1.51
<= 3000	2.42	3.08	2.16

Bi-Directional

	Expected	PMN.Aph	PMN
<= 2500	0.03	0.01	0.02
<= 5000	0.07	0.08	0.04

Infection of human neutrophils with *A. phagocytophilum* increases DNA methylation upstream of promoters and far downstream of transcription termination sites. New regions of DNA methylation enrichment were investigated using CEAS to determine the percentage that were present in A) promoters, B) downstream of genes, and C) in bi-directional promoters. Only the most significant peaks, those with a $-10*\log_{10}(\text{pval}) > 100$, were analyzed. White bars denote the percentage of the genome annotated to that particular feature (genomic background), gray bars denote regions of DNA methylation unique to *A. phagocytophilum* infected human peripheral blood neutrophils, and black bars denote regions of methylation unique to uninfected cells. Tables in D shows the exact percentage of DNA methylation enrichment for each particular feature where PMN.Aph represents regions unique to infected cells, and PMN to regions unique to uninfected cells.

Figure 4.5

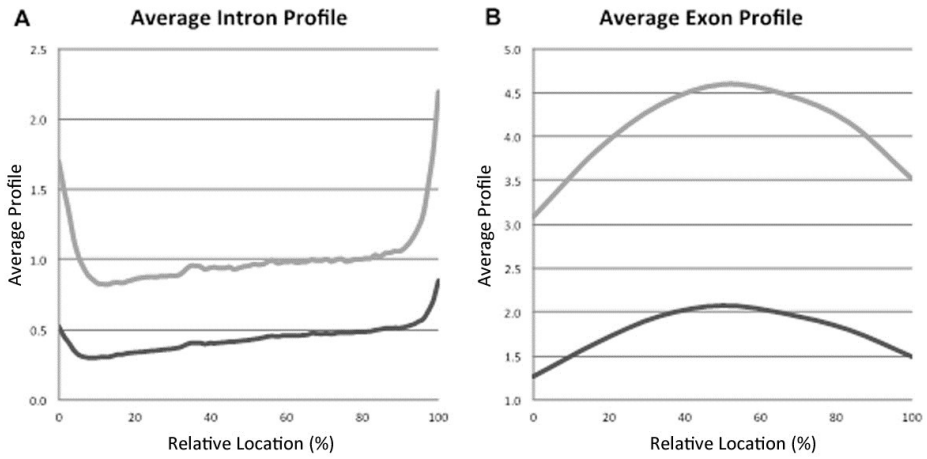


C

	Expected	PMN.Aph	PMN
5'UTR	0.44	0.38	0.30
3'UTR	1.48	1.40	0.93
Exon	1.93	5.04	3.68
Intron	42.21	44.16	28.85
All	46.26	50.98	33.76

***A. phagocytophilum* infection causes an increase in exon DNA methylation.** Identification of gene features with DNA methylation enrichment unique to *A. phagocytophilum* infected neutrophils and uninfected neutrophils was performed using CEAS. Only the most significant peaks, those with a $-10*\log_{10}(\text{pval}) > 100$, were analyzed. Features investigated include A) 5' and 3' UTRs, and B) exons, intron, and the sum of all gene features. White bars denote the percentage of the genome annotated to that particular feature (genomic background), gray bars denote regions of DNA methylation unique to *A. phagocytophilum* infected human peripheral blood neutrophils, and black bars denote regions of methylation unique to uninfected cells. Tables in C shows the exact percentage of DNA methylation enrichment for each particular feature where PMN.Aph represents regions unique to infected cells, and PMN to regions unique to uninfected cells.

Figure 4.6



Average signal profile for introns and exons. The average signal profiles for A) medium-sized introns (842-2,715 bp) and B) exons (109-164 bp) was calculated using CEAS. Light gray represents the average enrichment profile for *A. phagocytophilum*-infected human peripheral blood neutrophils and dark gray represents uninfected neutrophils.

DNA methylation and transcriptional alterations in the *A. phagocytophilum*-infected neutrophil.

We then sought to determine whether or not DNA methylation influenced transcription of the most differentially expressed genes. Genes that were up- or down-regulated with a fold change greater than 3 or 6 standard deviations from the mean fold change were assessed for DNA methylation peaks. If a region of DNA methylation was located within the gene body or ±3 kb from TSS or TTS, respectively, it was considered associated with that gene. Using these criteria, we compared the fold change in expression of 516 genes associated with DNA methylation to the remaining genes (not associated with DNA methylation) (Figure 4.7A). As anticipated, genes with differential gene expression (DGE) greater than 3 SD from the mean that were associated with DNA methylation were downregulated ($p<0.001$). However, DGE deviating $\geq$ 6 SD from the mean were upregulated rather than downregulated, and the resulting association with DNA methylation was strongly positive ($p<10^{-5}$).

Next, we investigated the effect of DNA methylation on transcription by parsing locations of DNA methylation within gene features. Genes associated with DNA methylation were sorted based on the gene feature location of the methylation site. Gene features investigated included both 3' and 5' UTRs, coding sequences (CDS), intronic regions, intergenic regions (>3kb from TSS or TTS) and promoters (±3kb from TSS or TTS). Additionally, the fold change in genes associated with DNA methylation was compared to those without. Our findings indicate that intergenic DNA methylation had a statistically significant impact on nearby gene expression ($p<0.05$) but the overall expression fold change was low (Figure 4.7B). As anticipated, genes associated with

DNA methylation in any location were generally downregulated ($p<0.01$).

Although *A. phagocytophilum* infection induced genome hypermethylation, the native level of genome methylation in neutrophils was high. Paired with the relatively low fold changes in gene expression and the small proportion of all genes that were differentially regulated, these data could mask correlations that would provide evidence for the effects that *A. phagocytophilum* has on DNA methylation that in turn lead to altered gene expression. Because of this, we set out to identify local effects at the chromosomal level. Recently, we discovered that large regions of chromosomes were regulated similarly during infection, suggesting large scale host cell chromatin reorganization with *A. phagocytophilum* infection.[34] This discovery, in light of genome hypermethylation, further supported the hypothesis that epigenetic alterations coordinate infected granulocyte transcriptional reprogramming. The MHC locus on chromosome 6 was shown to have local clustering of increased gene expression across a large chromosomal area spanning nearly 4 Mb.[34] This locus has the added advantage of well-described epigenetic regulation; thus, we focused attention on the association of gene expression and DNA methylation here.

To examine whether increased DNA methylation at the MHC locus on chromosome 6 was related to the observed cluster of differentially regulated genes, we examined new meDNA fold-enrichment at individual genes and created a sliding scale window for both the $\log_2$ fold-change in expression and newly methylated sites in those windows.

When individual newly methylated sites across the entire chromosome were

examined, no correlation between differential gene expression and meDNA fold-enrichment (ρ -0.0474, p=0.295) or $-\log_{10}$ p value (ρ -0.0189 and p=0.677) could be identified. In contrast, over the MHC locus, differential gene expression had a significant positive association with newly methylated sites, fold-enrichment (ρ 0.324; p= 0.031) and $-\log_{10}$ p value (ρ 0.3572; p= 0.016), but not at the region immediately downstream (meDNA fold enrichment ρ -0.003, p= 0.996; $-\log_{10}$ p value ρ -0.0178 and p=0.948). To determine whether specific genomic loci over long ranges on chromosomes were significantly enriched for newly methylated DNA in areas where a correlation with differential gene expression could be demonstrated, we plotted "ρ", the Spearman correlation coefficient for fold-enrichment of new meDNA and differential gene expression over windows (median 4.1 Mb [IQR 3.9]; median 14 meDNA features [IQR 2]) against the physical chromosome position of each window on the p and q arms of chromosome 6. We also plotted the p-value (1-p) for the correlations at each window to better visualize genomic regions where DNA methylation could have a significant and broad impact on gene transcription (Figure 4.8). The plot demonstrated long genomic stretches of positive and negative correlation between fold enrichment in DNA methylation and gene expression in the windows, including several at which significant associations were found, such as at the MHC locus. To explore these long range genomic associations between new DNA methylation and differential gene expression, we examined these large scale windows on chromosome 6 and the MHC locus in more detail.

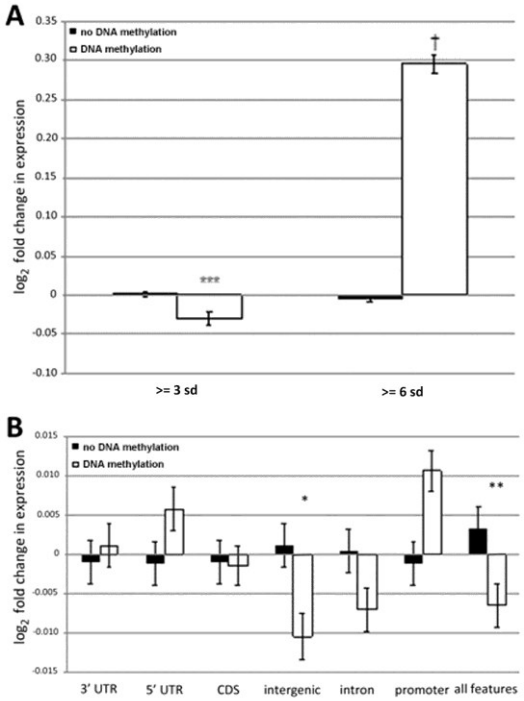
Across chromosome 6, as observed when individual differential gene expression and newly meDNA were examined, there was a slightly negative (ρ=-0.021) but non-

significant (p=0.5802) correlation of windows with newly methylated DNA fold-enrichment and $\log_2$ fold differential gene expression windows. In contrast, across the 3.1 Mb MHC region analyzed, there was a strongly positive correlation between window average new meDNA fold-increase and window average $\log_2$ fold-change for differential expression (ρ=0.422; p=0.002); for the similar sized (3.2 Mb) region immediately downstream from the MHC locus (PROX), the correlation was not significant (ρ=0.332; p=0.131) (Figure 4.9). When expression was sorted by upregulated vs. downregulated windows over the MHC locus, a significant positive correlation was observed for upregulated (ρ=0.447; p=0.003) but not downregulated conditions (ρ=0.133; p=0.683).Of interest, when windows were sorted by intragenic vs. intergenic features, there was a significant association of meDNA vs. differential regulation at the MHC locus for intragenic (ρ=0.644; p= 0.001) but not for intergenic features (ρ=0.144; p= 0.438) but the converse was true at PROX r (intergenic ρ=0.788; p= 0.002; intragenic ρ=0.042; p= 0.919) (Supplemental Figure 4.1).

To deconstruct this further, we subdivided individual DNA methylation peaks and differentially expressed genes over chromosome 6 into their gene regions: intron, CDS, intergenic and promoter. Here, no single gene region of chromosome 6 had new DNA methylation fold enrichment that significantly correlated, either positively or negatively, with differential gene transcription (Spearman $p \geq 0.091$), although the number of meDNA peaks in many categories was relatively small. Only 16 peaks for meDNA enrichment were found in gene promoters, yielding a Spearman ρ value of -0.438 (p= 0.091). Because those peaks were located outside of gene bodies, they were grouped with intergenic regions. This grouping increased the intergenic region Spearman coefficient to

-0.017, suggesting that although these 16 regions could influence expression of their respective genes, the likely impact across all genes on chromosome 6 is not high. Similar findings were obtained when the windows corresponding to these regions were evaluated, where the greatest correlation was in the 5' UTR regions, yet the ρ value was -0.223, with a p value of 0.247. The window regions corresponding to the MHC locus could not be evaluated in the same manner because of the paucity of data over windows in this region. Yet in specific genomic loci, the impact of newly meDNA regions is important, especially considering that newly meDNA was present in 47% of intergenic and promoter regions across chromosome 6, but in 58% of the intergenic regions in the MHC locus.

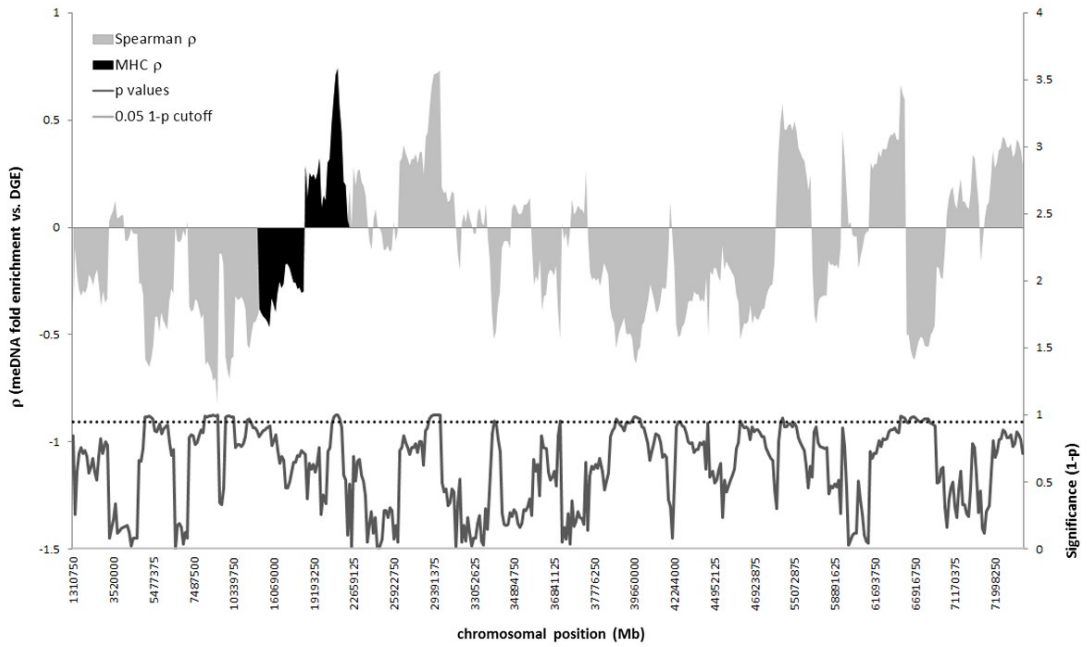
Figure 4.7



Gene Feature	# mDNA Peaks
3' UTR	354
5' UTR	644
CDS	254
Intergenic	4005
Intron	4027
Promoter	382
All Features	9666

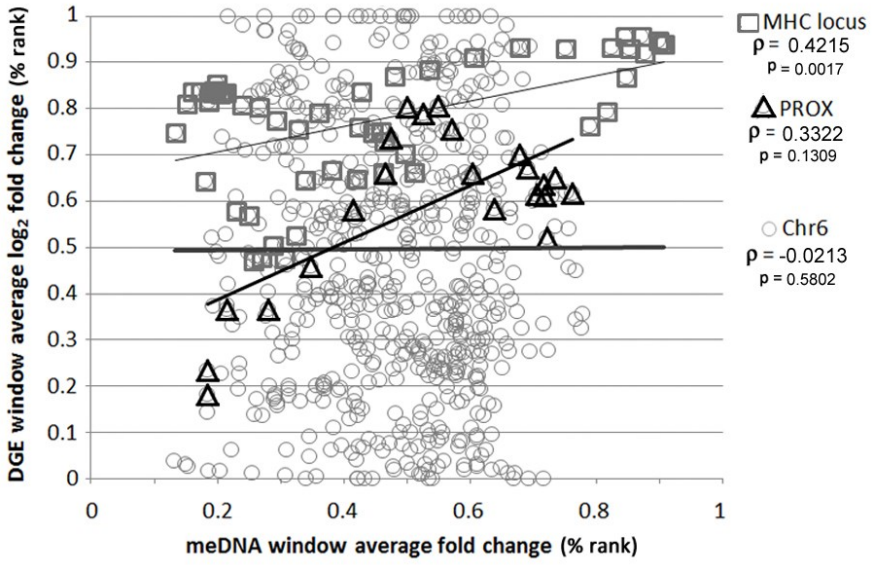
Effect of DNA methylation on gene regulation. The newly methylated regions that also had differential gene expression were sorted by standard deviation from the mean of differential gene expression to identify those most differentially expressed (A). Associated gene features were similarly sorted (B): 3'UTR, 5'UTR, coding sequences (CDS), intergenic regions, introns and promoters and correlated with $\log_2$ fold change, or differential gene expression, between infected and uninfected samples for their respective genes. DNA methylation peaks were considered intergenic when located 3kb or more from a TSS or TTS and effects on transcription were determined for the nearest gene. Transcriptional changes of genes with DNA methylation were compared to those without. *, $p<0.05$, **, $p<0.01$, ***, $p<0.001$, †, $p<10^{-4}$.

Figure 4.8



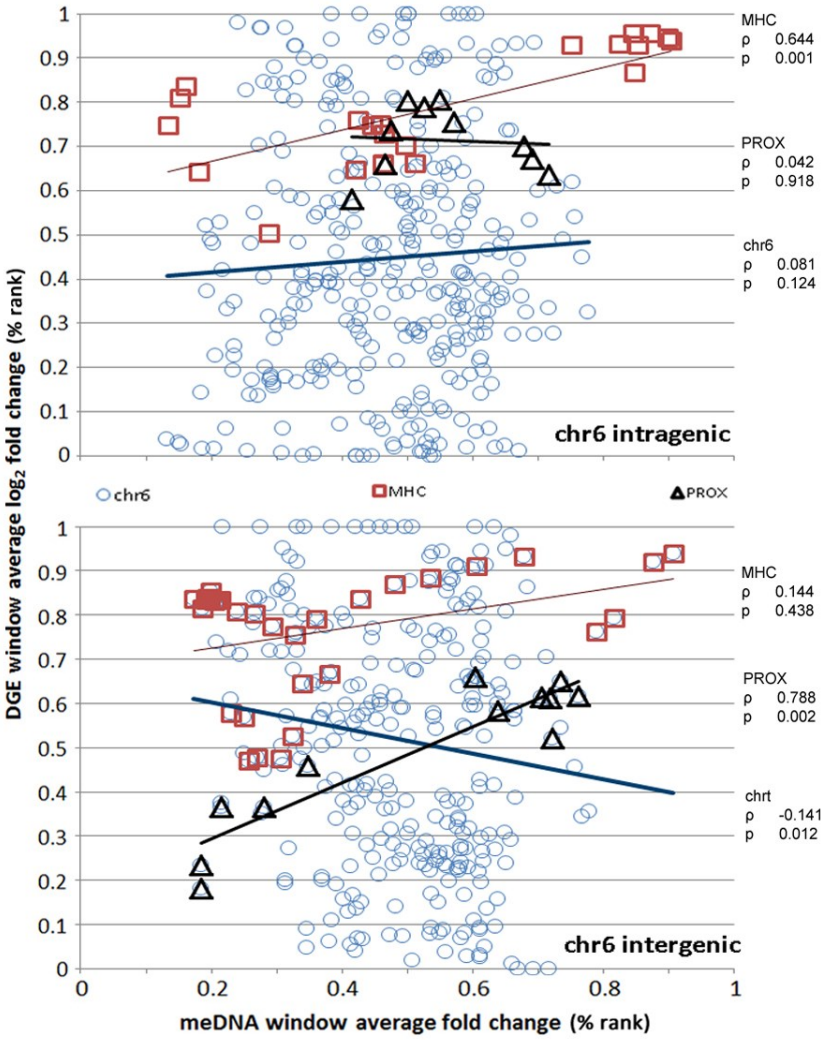
Large genomic regions on chromosome 6 demonstrate correlations between new DNA methylation fold-enrichment and differential gene expression after *A. phagocytophilum* infection of human peripheral blood neutrophils. Chromosome 6 positions are displayed in Mb across the x-axis and the Spearman correlation coefficient (ρ) between fold-enrichment of meDNA and $\log_2$ fold change in differential gene expression (DGE) is plotted on the left y-axis. Gray zones correspond to regions of chromosome 6 that are upregulated (positive) and downregulated (negative); black zones delineated the boundaries of the MHC locus. The lower black plot corresponds to the right y-axis and shows 1-p, for the Spearman correlation at the window centered at that genomic position. The hashed line delineates where the 1-p value is above 0.95 (equivalent to $p<0.05$).

Figure 4.9



Changes in meDNA with *A. phagocytophilum* infection correlate with regions of genomic differential expression. A) Increased DNA methylation is significantly correlated with differential gene expression across the MHC locus on chromosome 6, but not over the entire chromosome (Chr6) or across a similarly-sized genomic region immediately downstream of the MHC locus (PROX).

Supplemental Figure 4.1



Intragenic and intergenic DNA methylation correlate with differential gene expression within the MHC locus and downstream region. Increased DNA methylation within intragenic regions across the MHC locus correlate with differential gene expression. Conversely, intergenic DNA methylation downstream (PROX) of the MHC locus is correlated with differential gene expression.

Discussion

Alteration of DNA methylation patterns is well documented in stem cell differentiation, cellular reprogramming, cancers, and other disease states, but not in bacterial infections.[20,21,35-38] Previously, several groups suggested that infection by microbial pathogens induces hypermethylation of candidate genes involved in oncogenesis.[39-41] *Helicobacter pylori* infection causes chronic inflammation of surrounding gastric tissue resulting in aberrant DNA methylation patterns which can contribute to neoplastic transformation.[40] Uropathogenic *E. coli* can induce hypermethylation of *CDKN2A*, a cyclin-dependent kinase inhibitor and tumor suppressor gene, leading to increased proliferation of the urothelium.[41] Most recently, methylomes of ocular adnexal extranodal marginal zone B-cell lymphoma (EMZL) tumors from patients with *Chlamydophila psittaci* infection were compared to EMZL *C. psittaci*-negative tumors.[42] Differentially methylated regions of infected samples hierarchically clustered together suggesting changes in DNA methylation patterns were specific to *C. psittaci* infection.[42] Though informative, two of these studies investigated tissue samples which inevitably contain heterogeneous cell populations, and the other only investigated several candidate genes. Here, we show a direct link between infection of *A. phagocytophilum* and genome-wide alterations of DNA methylation profiles of the human neutrophil.

A. phagocytophilum infection induces widespread hypermethylation of the neutrophil genome within 24 hours of intracellular infection. Hypermethylation of the neutrophil genome benefits the bacterium and allows for increased colonization of the host cell since inhibition of DNMTs by 5-azacytidine resulted in decreased bacterial loads at 24 h independent of dose. This further suggests that even small changes in DNA

methylation affect bacterial growth. Previous studies investigating neutophil methylation patterns focused changes induced by myeloid differentiation and regions of hypomethylation.[37,38] Despite the hypermethylation of the host genome induced by *A. phagocytophilum*, many of the key regions identified as being hypomethylated remained unmethylated yet had reduced expression in infected cells, suggesting more than DNA methylation is at play.

Regions of methylation were found to localize within gene bodies and gene-associated DNA within 3 kb of the transcriptional start and termination sites. Intragenic DNA methylation is associated with regulation of alternative promoters, alternative splicing, and non-coding RNAs (ncRNAs).[43-47] Genes with intragenic DNA methylation continue to be expressed or have increased expression. It is thought that CpG methylation in intragenic regions results in decreased transcription from alternate promoters and thus, decreased transcription of regulating ncRNAs that in turn promotes increased target gene transcription.[44,47] In contrast, Jjingo *et al.* suggest that increased transcription and gene body methylation could be a result of DNA accessibility to methylating complexes, rather than regulation by ncRNA or expression of alternate promoters.[45] Previously published reports regarding neutrophil transcription in response to *A. phagocytophilum* infection show that more genes are upregulated than downregulated.[8-10] The chromosome 6 MHC locus was one of the regions with the most differentially expressed genes at 24 h in these datasets and these findings correlate with new regions of DNA methylation demonstrated here. Whether the result of ncRNAs, the effects of methylating complexes, or other mechanisms, hypermethylation of intragenic sequences could support marked changes in transcriptional programs, including those that benefit *A. phagocytophilum* survival in

neutrophils. Owing to the suboptimal infection levels obtained in the Borjesson et al. studies, the proximate cause of these changes will require further investigation.[34]

Closer examination of new methylation sites in infected samples shows higher than expected localization within exons and intron/exon junctions. Rönnerblad *et al* independently confirmed a large number of differentially methylated regions in human neutrophils localized after the first exon in gene bodies.[37] DNA methylation at intron/exon junctions is associated with increased H3K36 trimethylation resulting in alternative splice variants.[43] Currently it is not known whether DNA methylation is the precipitating factor for histone H3K36 methylation or vice versa. Given these findings, we hypothesize that deep RNA-seq of *A. phagocytophilum* infected neutrophils could identify splice variants of highly expressed genes. The localization of newly methylated regions in gene bodies suggests that transcription of these genes could be enhanced and that splice variants are likely to exist; these observations could shed light on some aspects of the transcriptional reprogramming that is key to *A. phagocytophilum* survival and transmission. Previous reports utilizing transcriptional profiling by microarrays[8,9] are unlikely to detect or quantify the majority of potential variants.

Despite the marked increase in gene-associated DNA methylation induced by *A. phagocytophilum*, there is not a strong genome-wide correlation between infection-induced alteration in gene transcription and annotated regions of meDNA. Intragenic DNA methylation seems to have no effect on expression on a genome-wide scale. This could suggest that most newly methylated regions were not transcriptionally active, and that the potential for upregulated expression is abrogated by *A. phagocytophilum* infection. We now have evidence to suggest intergenic DNA methylation influences

transcriptional programs of nearby genes as observed across the genome as a whole. Within the MHC locus on chromosome 6, regions significantly enriched for meDNA appear to demarcate boundaries of chromatin with increased expression and localize within regions of differential expression. This particular arrangement suggests that DNA methylation, and potentially histone marks, contribute to alterations over large chromosomal territories that influence expression of surrounding genes and accompanying transcriptional programs. Based on these observations, DNA methylation or factors that influence DNA methylation and DGE across large genomic regions could be important aspects of *A. phagocytophilum* neutrophil infection and thus, disease pathogenesis.

The conceptual framework regarding genomic regions of differential methylation has scarcely been examined. However, recent studies in colonic carcinoma reveal that focal regions, >100 kb in size, correlate with regions of attachment to the nuclear lamina, suggesting that factors that bring together distant coordinately regulated genomic regions, such as those tethered by matrix attachment region binding proteins, could guide DNA methylation, histone acetylation, and perhaps ncRNA expression to direct transcriptional reprogramming and pathogen survival.[48] Much remains to be determined, and an important question includes whether *A. phagocytophilum* plays a passive or active role in the process. A compelling hypothesis could revolve around the known translocation of the *A. phagocytophilum* nucleomodulin AnkA that binds to DNA in multiple genomic locations and recruits histone-modifying complexes including those known to interact with DNA methyltransferases and methylated CpG-binding domain proteins.[16,17] The specificity of AnkA-DNA binding could then dictate genome-wide epigenetic alterations

that reprogram cells to support *A. phagocytophilum* survival instead of microbial killing. Indeed, whether such reprogramming events occur as a result of other prokaryotic nucleomodulins is an area of intense research by investigators of microbial pathogenesis. The epigenetic research tools pioneered by oncologists, stem cell biologists, and those who study reprogramming of somatic cells now find applications in microbial pathogenesis studies as well. With mounting evidence of the impact of nucleomodulins from *Shigella, Listeria, Legionella, Chlamydia, Mycobacterium, Toxoplasma*, and other pathogenic microbes comes the realization that their impact surmounts the classical appreciation of influences over signaling pathways and intracellular trafficking, to the control of entire cellular programs.

Conclusion

Within 24h of infection, *A. phagocytophilum* induces hypermethylation of neutrophil DNA on a genome-wide scale. Hypermethylation of the genome is associated with enhanced bacterial growth and colonization of the host cell. DNA methylation was predominantly associated with gene bodies or within 3 kb of transcriptional start and termination sites, and tended to localize to exons and intron/exon junctions. Intragenic DNA methylation did not correlate strongly with *A. phagocytophilum*-induced differential gene expression over all chromosomes. However, intergenic DNA methylation is associated with infection-related transcriptional alterations. Targeted investigation of the MHC locus on chromosome 6 demonstrated that newly methylated sites fall within inactive chromatin domains and line the boundaries of active chromatin and that DNA methylation is associated with differential gene expression. Together, these data suggest that *A. phagocytophilum*-induced hypermethylation of the neutrophil genome is important for initial survival and propagation, and that its influence over cellular epigenetics can lead to cellular reprogramming of even terminally-differentiated neutrophils.

Acknowledgements. The authors wish to acknowledge the contributions and discussions with members of SHS's thesis committee, including Stephen Baylin, M.D. (Johns Hopkins University School of Medicine), Srinivasan Yegnasubramanian M.D., Ph.D. (Johns Hopkins University School of Medicine), and Frank DeLeo, Ph.D. (Rocky Mountain Laboratory, National Institute of Allergy and Infectious Disease). The authors would also like to thank Kristen Rennoll-Bankert, Ph.D. and Jose Carlos Garcia-Garcia, Ph.D. for their discussion and day-to-day support. Thank you to S. Yegnasubramanian,

M.D., Ph.D. and Sarah Wheelan, M.D., Ph.D. and the Johns Hopkins University Sidney Kimmel Comprehensive Cancer Center Next Generation Sequencing Center (http://nextgenseq.onc.jhmi.edu) for performing the MBD-seq and Conover Talbot, B.S. from the Johns Hopkins Deep Sequencing and Microarray Core for his assistance re-analyzing microarray using current tools and for the thoughtful discussion. Additional thanks to Cosme Adrover-Pacheco, Ph.D. for help with biocomputing analyses. Funding to support this work was provided by NIAID grant R01AI044102 to JSD. SY was supported by NIH/NCI grant P30CA006973, the Cleveland Foundation Masenhimer Fellowship.

References

1. Chen SM, Dumler JS, Bakken JS, Walker DH. Identification of a granulocytotropic ehrlichia species as the etiologic agent of human disease. *J Clin Microbiol*. 1994;32(3):589-595.

2. Carlyon JA, Chan WT, Galan J, Roos D, Fikrig E. Repression of *rac2* mRNA expression by *Anaplasma phagocytophila* is essential to the inhibition of superoxide production and bacterial proliferation. *J Immunol*. 2002;169(12):7009-7018.

3. Carlyon JA, Fikrig E. Mechanisms of evasion of neutrophil killing by *Anaplasma phagocytophilum*. *Curr Opin Hematol*. 2006;13(1):28-33.

4. Carlyon JA, Fikrig E. Invasion and survival strategies of *Anaplasma phagocytophilum*. *Cell Microbiol*. 2003;5(11):743-754.

5. Dumler JS, Choi KS, Garcia-Garcia JC, et al. Human granulocytic anaplasmosis and *Anaplasma phagocytophilum*. *Emerg Infect Dis*. 2005;11(12):1828-1834.

6. Choi KS, Park JT, Dumler JS. *Anaplasma phagocytophilum* delay of neutrophil apoptosis through the p38 mitogen-activated protein kinase signal pathway. *Infect Immun*. 2005;73(12):8209-8218.

7. Garyu JW, Choi KS, Grab DJ, Dumler JS. Defective phagocytosis in *Anaplasma phagocytophilum*-infected neutrophils. *Infect Immun*. 2005;73(2):1187-1190.

8. Borjesson DL, Kobayashi SD, Whitney AR, Voyich JM, Argue CM, Deleo FR. Insights into pathogen immune evasion mechanisms: *Anaplasma phagocytophilum* fails to induce an apoptosis differentiation program in human neutrophils. *J Immunol*. 2005;174(10):6364-6372.

9. de la Fuente J, Ayoubi P, Blouin EF, Almazan C, Naranjo V, Kocan KM. Gene expression profiling of human promyelocytic cells in response to infection with *Anaplasma phagocytophilum*. *Cell Microbiol*. 2005;7(4):549-559.

10. Lee HC, Kioi M, Han J, Puri RK, Goodman JL. *Anaplasma phagocytophilum*-induced gene expression in both human neutrophils and HL-60 cells. *Genomics*. 2008;92(3):144-151.

11. Banerjee R, Anguita J, Roos D, Fikrig E. Cutting edge: Infection by the agent of human granulocytic ehrlichiosis prevents the respiratory burst by down-regulating *gp91phox*. *J Immunol*. 2000;164(8):3946-3949.

12. Ge Y, Yoshiie K, Kuribayashi F, Lin M, Rikihisa Y. *Anaplasma phagocytophilum* inhibits human neutrophil apoptosis via upregulation of *bfl-1*, maintenance of mitochondrial membrane potential and prevention of caspase 3 activation. *Cell Microbiol*. 2005;7(1):29-38.

13. Akkoyunlu M, Malawista SE, Anguita J, Fikrig E. Exploitation of interleukin-8-induced neutrophil chemotaxis by the agent of human granulocytic ehrlichiosis. *Infect Immun*. 2001;69(9):5577-5588.

14. Klein MB, Hu S, Chao CC, Goodman JL. The agent of human granulocytic ehrlichiosis induces the production of myelosuppressing chemokines without induction of proinflammatory cytokines. *J Infect Dis*. 2000;182(1):200-205.

15. Scorpio DG, Akkoyunlu M, Fikrig E, Dumler JS. CXCR2 blockade influences *Anaplasma phagocytophilum* propagation but not histopathology in the mouse model of human granulocytic anaplasmosis. *Clin Diagn Lab Immunol*. 2004;11(5):963-968.

16. Garcia-Garcia JC, Barat NC, Trembley SJ, Dumler JS. Epigenetic silencing of host cell defense genes enhances intracellular survival of the rickettsial pathogen *Anaplasma phagocytophilum*. *PLoS Pathog*. 2009;5(6):e1000488.

17. Garcia-Garcia JC, Rennoll-Bankert KE, Pelly S, Milstone AM, Dumler JS. Silencing of host cell CYBB gene expression by the nuclear effector AnkA of the intracellular pathogen *Anaplasma phagocytophilum*. *Infect Immun*. 2009;77(6):2385-2391.

18. Stirzaker C, Song JZ, Davidson B, Clark SJ. Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. *Cancer Res*. 2004;64(11):3871-3877.

19. Srivastava S, Mishra RK, Dhawan J. Regulation of cellular chromatin state: Insights from quiescence and differentiation. *Organogenesis*. 2010;6(1):37-47.

20. Baylin SB, Herman JG. DNA hypermethylation in tumorigenesis: Epigenetics joins genetics. *Trends Genet*. 2000;16(4):168-174.

21. Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum Mol Genet*. 2001;10(7):687-692.

22. Rennoll-Bankert KE, Sinclair SH, Lichay MA, Dumler JS. Comparison and characterization of granulocyte cell models for *Anaplasma phagocytophilum* infection. *Pathog Dis*. 2013.

23. Goodman JL, Nelson C, Vitale B, et al. Direct cultivation of the causative agent of human granulocytic ehrlichiosis. *N Engl J Med*. 1996;334(4):209-215.

24. Guerrero-Preston R, Michailidi C, Marchionni L, et al. Key tumor suppressor genes inactivated by "greater promoter" methylation and somatic mutations in head and neck cancer. *Epigenetics*. 2014;9(7).

25. Aryee MJ, Liu W, Engelmann JC, et al. DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Sci Transl Med*. 2013;5(169):169ra10.

26. Yegnasubramanian S, Lin X, Haffner MC, DeMarzo AM, Nelson WG. Combination of methylated-DNA precipitation and methylation-sensitive restriction enzymes (COMPARE-MS) for the rapid, sensitive and quantitative detection of DNA methylation. *Nucleic Acids Res*. 2006;34(3):e19.

27. Yegnasubramanian S, Wu Z, Haffner MC, et al. Chromosome-wide mapping of DNA methylation patterns in normal and malignant prostate cells reveals pervasive methylation of gene-associated and conserved intergenic sequences. *BMC Genomics*. 2011;12:313-2164-12-313.

28. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol*. 2008;9(9):R137-2008-9-9-r137. Epub 2008 Sep 17.

29. Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. A clustering approach for identification of enriched domains from histone modification ChIP-seq data. *Bioinformatics*. 2009;25(15):1952-1958.

30. Easwaran H, Johnstone SE, Van Neste L, et al. A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res*. 2012;22(5):837-849.

31. Karimi M, Johansson S, Stach D, et al. LUMA (LUminometric methylation assay)--a high throughput method to the analysis of genomic DNA methylation. *Exp Cell Res*. 2006;312(11):1989-1995.

32. Ji X, Li W, Song J, Wei L, Liu XS. CEAS: Cis-regulatory element annotation system. *Nucleic Acids Res*. 2006;34(Web Server issue):W551-4.

33. Shin H, Liu T, Manrai AK, Liu XS. CEAS: Cis-regulatory element annotation system. *Bioinformatics*. 2009;25(19):2605-2606.

34. Sinclair SH, Rennoll-Bankert KE, Dumler JS. Effector bottleneck: Microbial reprogramming of parasitized host cell transcription by epigenetic remodeling of chromatin structure. *Front Genet*. 2014;5:274.

35. Ciechomska M, van Laar JM, O'Reilly S. Emerging role of epigenetics in systemic sclerosis pathogenesis. *Genes Immun*. 2014.

36. de Mello VD, Pulkkinen L, Lalli M, Kolehmainen M, Pihlajamaki J, Uusitupa M. DNA methylation in obesity and type 2 diabetes. *Ann Med*. 2014;46(3):103-113.

37. Ronnerblad M, Andersson R, Olofsson T, et al. Analysis of the DNA methylome and transcriptome in granulopoiesis reveals timed changes and dynamic enhancer methylation. *Blood*. 2014;123(17):e79-89.

38. Zilbauer M, Rayner TF, Clark C, et al. Genome-wide methylation analyses of primary human leukocyte subsets identifies functionally important cell-type-specific hypomethylated regions. *Blood*. 2013;122(25):e52-60.

39. Kondo T, Oka T, Sato H, et al. Accumulation of aberrant CpG hypermethylation by *Helicobacter pylori* infection promotes development and progression of gastric MALT lymphoma. *Int J Oncol*. 2009;35(3):547-557.

40. Matsusaka K, Funata S, Fukayama M, Kaneda A. DNA methylation in gastric cancer, related to *Helicobacter pylori* and Epstein-Barr virus. *World J Gastroenterol*. 2014;20(14):3916-3926.

41. Tolg C, Bagli DJ. Uropathogenic *Escherichia coli* infection: Potential importance of epigenetics. *Epigenomics*. 2012;4(2):229-235.

42. Lee MJ, Min BJ, Choung HK, Kim N, Kim YA, Khwarg SI. Genome-wide DNA methylation profiles according to *Chlamydophila psittaci* infection and the response to doxycycline treatment in ocular adnexal lymphoma. *Mol Vis*. 2014;20:1037-1047.

43. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. Epigenetics in alternative pre-mRNA splicing. *Cell*. 2011;144(1):16-26.

44. Maunakea AK, Nagarajan RP, Bilenky M, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. 2010;466(7303):253-257.

45. Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK. On the presence and role of human gene-body DNA methylation. *Oncotarget*. 2012;3(4):462-474.

46. Shenker N, Flanagan JM. Intragenic DNA methylation: Implications of this epigenetic mechanism for cancer research. *Br J Cancer*. 2012;106(2):248-253.

47. Kulis M, Queiros AC, Beekman R, Martin-Subero JI. Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochim Biophys Acta*. 2013;1829(11):1161-1174.

48. Berman BP, Weisenberger DJ, Aman JF, et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet*. 2011;44(1):40-46.

Chapter 5: Bioinformatic and mass spectrometry identification of *Anaplasma phagocytophilum* proteins translocated into host cell nuclei

This work has been published in *Frontiers in Microbiology* and is reprinted here with permission from Frontiers Media.

Sinclair SH, Garcia-Garcia JC, Dumler JS. 2015. Bioinformatic and mass spectrometry identification of *Anaplasma phagocytophilum* proteins translocated into host cell nuclei. *Front Microbiol.* Feb 6(55):1-10/doi:10.3389/fmicb.2015.00055

Abstract

Obligate intracellular bacteria have an arsenal of proteins that alter host cells to establish and maintain a hospitable environment for replication. *Anaplasma phagocytophilum* secrets Ankyrin A (AnkA), via a type IV secretion system, which translocates to the nucleus of its host cell, human neutrophils. *A. phagocytophilum*-infected neutrophils have dramatically altered phenotypes in part explained by AnkA-induced transcriptional alterations. However, it is unlikely that AnkA is the sole effector to account for infection-induced transcriptional changes. We developed a simple method combining bioinformatics and iTRAQ protein profiling to identify potential bacterial-derived nuclear-translocated proteins that could impact transcriptional programming in host cells. This approach identified 50 *A. phagocytophilum* candidate genes or proteins. The encoding genes were cloned to create GFP fusion protein-expressing clones that were transfected into HEK-293T cells. We confirmed nuclear translocation of six proteins: APH_0062, RplE, Hup, APH_0382, APH_0385, and APH_0455. Of the six, APH_0455 was identified as a type IV secretion substrate and is now under investigation as a potential nucleomodulin. Additionally, application of this approach to other obligate intracellular bacteria such as *Mycobacterium tuberculosis, Chlamydia trachomatis* and other intracellular bacteria identified multiple candidate genes to be investigated.

Introduction

Anaplasma phagocytophilum is an obligate intracellular bacterium of human neutrophils. The neutrophil is an unlikely host as it creates an intracellular milieu that is a highly inhospitable environment for bacterial survival. Yet, *A. phagocytophilum* requires the neutrophil for propagation and survives by altering the cellular antimicrobial properties while paradoxically increasing pro-inflammatory functions.[1-5] The fitness advantage gained with suppression of microbial killing while enhancing recruitment of new host cells for population expansion is the benefit of this paradoxical dichotomy of functional reprogramming. There is increasing evidence to suggest that the bacterium accomplishes this with coordinated reprogramming of neutrophil gene transcription by reorganizing large regions of host cell chromatin.[6]

Importantly, *A. phagocytophilum* produces a protein, Ankyrin A (AnkA) that is exported from the bacterium and eventually localizes to the nucleus of the infected host cell.[7,8] Previously, our laboratory investigated the effect of infection on the transcriptional repression of *CYBB*, encoding gp91^{phox}.[9,10] AnkA is capable of directly binding host cell DNA, and in the case of *CYBB*, transcription is dampened when AnkA binds to its proximal promoter.[8-10] Furthermore, increased histone deacetylase (HDAC) activity enhances *A. phagocytophilum* infection in part because AnkA recruits HDAC1 to the *CYBB* promoter to close the chromatin and exclude RNA polymerase binding.[10,11] Owing to their capacity to enter the nucleus and modulate host cell transcription, microbial factors such as AnkA have been called "nucleomodulins."

It is currently unclear as to whether HDAC recruitment is the predominant mechanism by which AnkA exerts its chromatin modulating effects, whether there are

other host factors (e.g. polycomb repressive or hematopoietic associated factor-1 [HAF1] complexes), or additional bacterial-derived nucleomodulins that further contribute to reprogramming. The *A. phagocytophilum* genome encodes a type 4 secretion system (T4SS) that allows the bacteria to translocate effector proteins into the host cytosol.[12-14] AnkA was the first T4SS substrate identified among Rickettsiales, and it plays a critical and potentially dominant role in the course of establishing and sustaining neutrophil infection.[9-11,15,16] In contrast to other gram-negative T4SSs, the *vir* genes encoding the secretion system of the *Rickettsiales* family are organized differently in that they are clustered in three different genomic locations.[13,17] Between individual *A. phagocytophilum* strains, variations of the T4SS appear to contribute to host specificity and strain virulence.[15]

We hypothesize that *A. phagocytophilum* expresses additional nuclear effector proteins secreted by its type IV secretion system (T4SS) and that these also play a role in pathogenicity. It is likely that some will be nucleomodulins which could contribute to transcriptional reprogramming of infected neutrophils. Pilot studies using bioinformatics tools, and iTRAQ protein profiling among infected and uninfected cells were used to identify candidate proteins that potentially localize to the host cell nucleus. The profiling identified 50 *A. phagocytophilum* proteins, one of which was AnkA, and at least 7 of these were predicted to enter the nucleus based on the presence of both a nuclear localization sequence and a bacterial secretion signal sequence. Ultimately, 3 of the 7 proteins identified in the bioinformatic screen and 3 of 37 identified by iTRAQ profiling of nuclei from infected cells translocated into HEK-293T human embryonic kidney and PLB-985 granulocytic cell nuclei.

Methods

In silico prediction of A. phagocytophilum proteins targeted to the host cell nucleus

Our initial focus was on proteins involved in regulation of host gene expression. Since these events occur chiefly in the nucleus, we developed an unbiased computational approach to identify potential nucleomodulins encoded in intracellular bacterial genomes based on their likelihood for translocation into the host cell nucleus and applied this to the *A. phagocytophilum* HZ strain genome (Supplemental Figure 5.1). Annotated protein tables for bacteria, focusing on the *A. phagocytophilum* HZ strain genome, were obtained from the National Center for Biotechnology Information (NCBI) database (ftp://ftp.ncbi.nih.gov/genomes/Bacteria). The *A. phagocytophilum* protein table was used as the database for eukaryotic subcellular localization search algorithms. Although we used a database with 1,264 annotated *A. phagocytophilum* proteins, including hypothetical proteins, multiple programs were implemented to obtain high prediction accuracy and processing capacity. Since we needed only to predict nuclear proteins, localization coverage was not taken into account. Since hybrid methods are preferable when little is known about the protein of interest [18] we used ProtComp Version 6 (Softberry, Inc.), a computational algorithm for the identification of sub-cellular localization of eukaryotic proteins.

We next applied PSORTb v.2.0 to exclude potential membrane proteins that are unlikely to be secreted into the host cell.[19] Finally, we used computational algorithms to predict the presence of eukaryotic nuclear localization signals (NLS). NLSs often possess sequences with a high basic amino-acid content [20] and are generally classified into three categories: classical or monopartite (NLSm), bipartite (NLSb), and a type of N-terminal

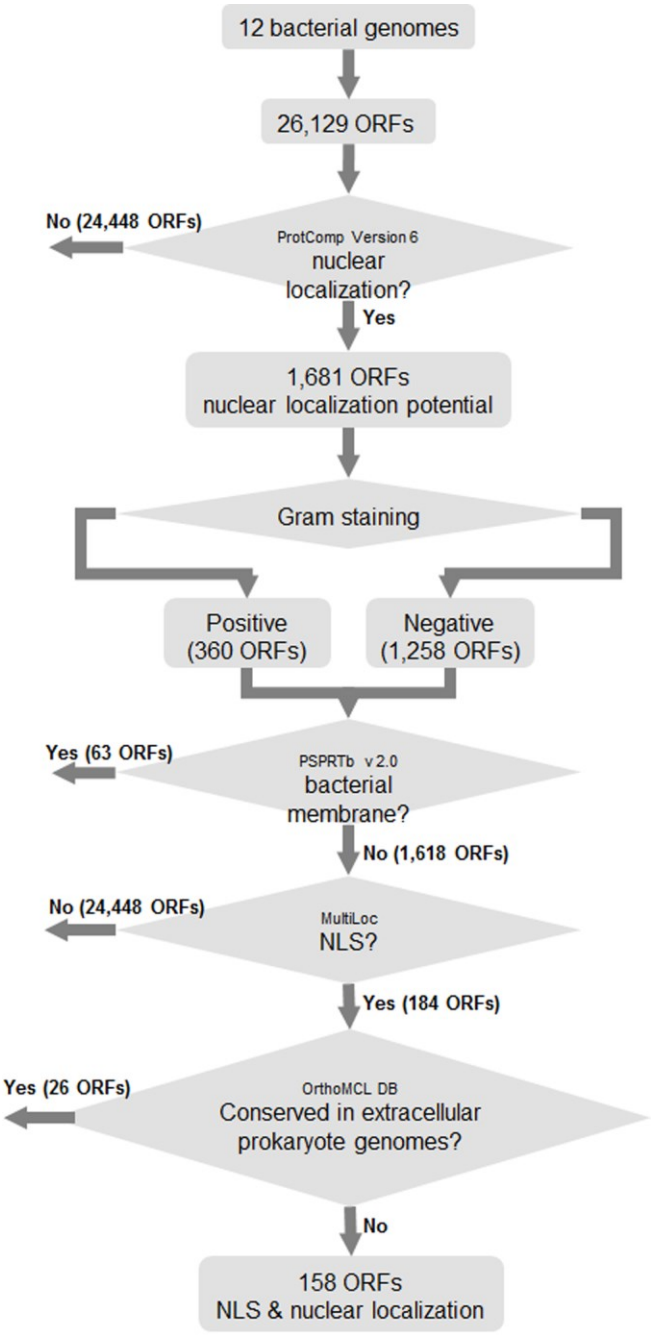
signal found in yeast protein, Mat alpha2, a poorly studied signal that is not incorporated in most NLS prediction algorithms. To screen broadly for potential NLSs, we selected MultiLoc.[21] MultiLoc also identifies matches in NLSdb, a database of experimentally known NLSs [22] and is also useful to also predict NLSm and NLSb in addition to the NLSdb attribute, since NLSdb recognizes only 43% of the nuclear proteins. MultiLoc calculates a probability estimate for each subcellular location and the protein is assigned to the compartment with the highest score. The MultiLoc output was recorded and used to calculate a Nuclear Score that better reflected the purpose of the search:

$$\textbf{Nuclear Score} = \textbf{MultiLoc Nuc} + \textbf{NLSdb} + (\mathbf{0.5} \times \textbf{NLSm} + \mathbf{0.5} \times \textbf{NLSb})$$

where: i) $0 \leq$ MultiLoc Nuc < 1 is the probability estimate of the protein being nuclear as calculated by MultiLoc; ii) NLSdb is 1 if the protein contains a known NLS, 0 if not; iii) NLSm is 1 if the protein contains a predicted NLSm, 0 if not; and iv) NLSb is 1 if the protein contains a predicted NLSb, 0 if not. Weighting was applied since the presence of a predicted NLS suggests, but is not conclusive; therefore the NLSm or NLSb prediction contributes only half to the final nuclear score. The addition of the continuous MultiLoc Nuc score provides a better ranking of the proteins, given that the other indicators contribute discretely to the Nuclear Score. However, no proteins without a known or predicted NLS will produce a Nuclear Score >1 since MultiLoc Nuc < 1. Finally, to exclude genes that are orthologs in extracellular bacteria and less likely to be important for intracellular survival, candidates of the algorithm were examined by OrthoMCL DB [23] while the concise microbial protein BLAST was used to determine whether selected ORFs are conserved in at least two of the intracellular bacterial genomes included in this study. OrthoMCL is a genome-scale algorithm for ortholog protein sequences which

identifies ORFs shared by two or more species/genomes. OrthoMCL starts by looking for reciprocal best hits within each genome as putative ortholog pairs. In a preliminary analysis, a cutoff E value < 10^{-100} was used to identify orthologs of candidate ORFs in extracellular bacteria genomes, using the *E. coli* genome as a reference.

Supplemental Figure 5.1



Algorithm used to discern potential nuclear translocated proteins in *A. phagocytophilum* HZ genome and among 11 additional genomes of intracellular bacteria. See methods for detailed description of programs and algorithm decision points.

***iTRAQ* for identification of potential nuclear-translocated protein profiling**

A. phagocytophilum-infected and uninfected HL-60 cells, a promyelocytic cell line commonly used for *A. phagocytophilum* propagation as previously described, were fractionated to obtain nuclei and nuclear proteins.[8] iTRAQ (isobaric tag for relative and absolute quantitation protein profiling technology [Applied Biosystems]), a mass spectrometric technique where 2 protein expression profiles are compared, was used to identify candidate bacterial proteins present in the nucleus of infected cells. One hundred µg in replicate samples from nuclear fractions of infected and uninfected HL-60 cells were acetone-precipitated and checked for protein integrity and sample quality. The samples were reduced and cysteines blocked following the iTRAQ kit protocol (Applied Biosystems). Samples were digested with trypsin overnight at 37°C and then labeled with iTRAQ tags in replicates, pooled and fractionated using a strong cation exchange (SCX) column on an Ultimate HPLC system (LC Packings, Netherlands). Approximately 20 fractions were collected and analyzed on Qstar Pulsar™ (Applied Biosystems-MDS Sciex) interfaced with an Agilent 1100 HPLC system. Peptides were separated on a reverse-phase column, and MS/MS analysis was performed. The MS/MS spectral data were extracted and searched against Uniprot-sprot database (entries for *Homo sapiens* and *A. phagocytophilum*) using ProteinPilot™ software (Applied Biosystems). For each protein, two types of scores were reported: unused ProtScore and total ProtScore. The total ProtScore is a measurement of all the peptide evidence for a protein and is analogous to protein scores reported by other protein identification software programs. However, the unused ProtScore is a measurement of all the peptide evidence for a protein that is not better explained by a higher ranking protein and was the method of choice. The

protein confidence threshold cutoff for this study was set at an unused score of 2.0 with at least one peptide with 99% confidence. A ratio of infected to uninfected (Aph:HL-60) score was used to identify *A. phagocytophilum* proteins in nuclear lysates. To do this, we averaged the ratios of uninfected HL-60 nuclear lysate replicates (isobaric isotope labels 115:114) and ratios of nuclear lysates replicates from *A. phagocytophilum*-infected HL-60 cells (116:114 and 117:114) to create the composite Aph:HL-60 mean ratio. Proteins identified with mean ratios (infected/uninfected) >1.2 were selected for further study.

GFP-fusion protein plasmid clones and transfections

41 GFP C-terminal fusion proteins were prepared using pMAXFP-Green-C (Lonza, cat# AMA-VDF1011) and the Infusion HD Liquid cloning kit (Clontech). Briefly, target genes were amplified using PlatinumTaq (Life Technologies) and PCR purified using Qiagen PCR purification kits (Qiagen). Primers were designed using Clontech's Online Infusion tools, Primer Design. Amplicons were created to be fused with the pMAXFP-Green-C vector after digestion with *Xho*I. Primers were approximately 40-45 bp in length and had the sequence GAAGAAAGATCTCGAGCT added to the 5' end of the forward gene-specific primer (20-25bp), and GAAGCTTGAGCTCGAGT added 5' to the reverse primer (Supplemental Table 5.1). The Infusion-HD kit instructions were followed as per the manufacturer's recommendations. Clones were transformed into *E. coli* JM109 (Promega) and after antibiotic selection, were sequenced to ensure they were in the correct orientation and in frame. HEK-293T cells were transfected with GFP-fusion vectors using Lipofectamine 2000 (Life Technologies) and PLB-985 cells were transfected with the Amaxa Nucleofector shuttle and the SF kit reagents (Lonza) as per manufacturer's

recommendations. PLB-985 cells are human myelomonoblast leukemia cells that easily differentiate into neutrophil-like cells and are readily transfected as opposed to HL-60 cells, a common host cell model for *A. phagocytophilum* infection. [24-26] Cells were stained with DAPI 24 h later and imaged by fluorescence microscopy, gathering both superimposed green fluorescent protein and DAPI images.

Supplemental Table 5.1. Primers used for amplification and cloning of potential *A. phagocytophilum* nuclear-translocated protein-encoding genes.

oligo name	orientation	sequence (5' to 3')
APH0062	forward	GAAGAAAGATCTCGAGCTATGCCTAGGGCGGTTGCCGTG
APH0062	reverse	GAAGCTTGAGCTCGAGTCACACCTTCTGGCGCTGGAGG
APH0097	forward	GAAGAAAGATCTCGAGCTATGCAGCCAAAAATTAGAGTAAGA
APH0097	reverse	GAAGCTTGAGCTCGAGTTTATTTCTCACCATCCGCATTC
APH0106	forward	GAAGAAAGATCTCGAGCTATGTTCAACGGCATAGTTACTC
APH0106	reverse	GAAGCTTGAGCTCGAGTCTATAGGCCATATGTGCGATG
APH0135	forward	GAAGAAAGATCTCGAGCTATGTCTCAGTTAGTGGATGGGAAGGCG
APH0135	reverse	GAAGCTTGAGCTCGAGTTACTCCGAAAGCACCTCTAGGTTCACAGC
APH0154	forward	GAAGAAAGATCTCGAGCTATGGTTGGGTACATTGGGGA
APH0154	reverse	GAAGCTTGAGCTCGAGTCTAGCAAGCCAAAGTATCGTAC
APH0240	forward	GAAGAAAGATCTCGAGCTATGTCAAATACGGTAGTCACG
APH0240	reverse	GAAGCTTGAGCTCGAGTTTAAAATCCGCCCATACCAC
APH0280	forward	GAAGAAAGATCTCGAGCTATGTGGATGAAAAAGGTTGGTA
APH0280	reverse	GAAGCTTGAGCTCGAGTTCAACCAGCCATGCTCTC
APH0288	forward	ATGAGTACAGTCGCTGAAATGGAAGAAAGATCTCGAGCT
APH0288	reverse	GAAGCTTGAGCTCGAGTTTAGACATTCTCTCCCCTTATCTT
APH0292	forward	GAAGAAAGATCTCGAGCTATGTTGGGGAGTTTATCTAAGGA
APH0292	reverse	GAAGCTTGAGCTCGAGTCTAATTCATAAAAGGAAATCCCAACT
APH0303	forward	GAAGAAAGATCTCGAGCTATGGCTGACCATTGGAACAA
APH0303	reverse	GAAGCTTGAGCTCGAGTTTAATCTTCACTATACTGCTTACTCAA
APH0339	forward	GAAGAAAGATCTCGAGCTATGAAGCAATATAAGTTATTGGAGCAA
APH0339	reverse	GAAGCTTGAGCTCGAGTTCACATGCTTAGATATCTATTTACCAG
APH0382	forward	GAAGAAAGATCTCGAGCTATGCATACGCCGCATATATTC
APH0382	reverse	GAAGCTTGAGCTCGAGTCTATGCGCTTCGCGCT
APH0385	forward	GAAGAAAGATCTCGAGCTATGCATACGCCTCGTATATTC
APH0385	reverse	GAAGCTTGAGCTCGAGTTTATGCACTTCGTGATCGTTT
APH0397	forward	ATGGGGAGTATACCAGAGTTTAGGAAGAAAGATCTCGAGCT
APH0397	reverse	GAAGCTTGAGCTCGAGTCTATTCTTGCACAACGGCAT
APH0398	forward	GAAGAAAGATCTCGAGCTATGAAAATTGATGTTGAGGTTATCAA
APH0398	reverse	GAAGCTTGAGCTCGAGTCTAAGCAATTGCGAATAACCTATAC
APH0445	forward	GAAGAAAGATCTCGAGCTATGGTGAATTTGTATAATTTTGATAATCTTG
APH0445	reverse	GAAGCTTGAGCTCGAGTCTAAACCCACCCAACTCGTT
APH0455	forward	GAAGAAAGATCTCGAGCTATGCATATGCCTCGTATATTCAC
APH0455	reverse	GAAGCTTGAGCTCGAGTTCATGCTCTTCGCGCCT
APH0469	forward	GAAGAAAGATCTCGAGCTATGGAAAACTCCCTACTGAAGAT
APH0469	reverse	GAAGCTTGAGCTCGAGTTCAACTTTTCTTCAAAAGAGTAGCT
APH0485	forward	GAAGAAAGATCTCGAGCTATGGAAAAAACTGCTATAGGTGAT
APH0485	reverse	GAAGCTTGAGCTCGAGTTTATATGAGATTACAGCTTTCGGTG
APH0515	forward	GAAGAAAGATCTCGAGCTATGAGAGAAAAACGTCGTCGT
APH0515	reverse	GAAGCTTGAGCTCGAGTTTACTTGCTATCTTTAGAGATAACCTT
APH0576	forward	GAAGAAAGATCTCGAGCTATGCGGGATTCATTGATGGAT
APH0576	reverse	GAAGCTTGAGCTCGAGTTTAGAAAAACCCACGCAATTTC
APH0629	forward	GAAGAAAGATCTCGAGCTATGCGGTCTCATCGTTCA
APH0629	reverse	GAAGCTTGAGCTCGAGTCTACTTTTGCTCAGAAATAATCTTCA
APH0659	forward	GAAGAAAGATCTCGAGCTATGAGACGCCTTATGAATTTGGT
APH0659	reverse	GAAGCTTGAGCTCGAGTTTAGCTCAGACTATCTATATGAGATTG
APH0740	forward	GAAGAAAGATCTCGAGCTATGTTGACAGAAGAAGAAAAGAAAAA

APH0740	reverse	GAAGCTTGAGCTCGAGTCTACCTACCGCGACCTC
APH0784	forward	GAAGAAAGATCTCGAGCTATGAGCAAGGAAATCATAGTAAGG
APH0784	reverse	GAAGCTTGAGCTCGAGTCTACACATTTACCGTATCTAACAAC
APH0805	forward	GAAGAAAGATCTCGAGCTATGAGGAAGTGGGTTATAAACTTTT
APH0805	reverse	GAAGCTTGAGCTCGAGTTCAGGCAATACTCTTTGAAGTG
APH0820	forward	GAAGAAAGATCTCGAGCTATGCTTACACCTAGGAATCTTAGT
APH0820	reverse	GAAGCTTGAGCTCGAGTTTAATGTAATCTGCGTCGTTTGG
APH0847	forward	GAAGAAAGATCTCGAGCTATGCATAATCATGGAAATCCATTAGGT
APH0847	reverse	GAAGCTTGAGCTCGAGTTTATTGTAAATTTGGGAATAACGTACTG
APH0906	forward	GAAGAAAGATCTCGAGCTATGACTCTGCTGCTTAAGCAA
APH0906	reverse	GAAGCTTGAGCTCGAGTTTAATGCTGCTGCTGTGATAC
APH0968	forward	GAAGAAAGATCTCGAGCTATGGAAGAGGGTAAAGTTGTGTT
APH0968	reverse	GAAGCTTGAGCTCGAGTTTAATTCAGAAAAGAACAATCACCAA
APH0971	forward	GAAGAAAGATCTCGAGCTATGCAGCAGTTCTATTGTGTC
APH0971	reverse	GAAGCTTGAGCTCGAGTCTACACACTCTCAAAAAGCGAT
APH1023	forward	GAAGAAAGATCTCGAGCTATGAAGACGTTGGATTTGTATGG
APH1023	reverse	GAAGCTTGAGCTCGAGTCTAGTAATCCACAACGGAGC
APH1025	forward	GAAGAAAGATCTCGAGCTATGAGTAGTGTTGATTTAGATGATTTG
APH1025	reverse	GAAGCTTGAGCTCGAGTTTACTTCAAACTCACCTGAGC
APH1027	forward	GAAGAAAGATCTCGAGCTATGTTGGATAAGAATATTGTCTATACTCCA
APH1027	reverse	GAAGCTTGAGCTCGAGTTCAACCCAATCCAGTTATTCTAAA
APH1029	forward	GAAGAAAGATCTCGAGCTATGAAGCATGATGGTATTGGTC
APH1029	reverse	GAAGCTTGAGCTCGAGTTTATTCCCCTACCTTCTGTATACTAC
APH1032	forward	GAAGAAAGATCTCGAGCTATGACAGAAGGAAGGAAGCC
APH1032	reverse	GAAGCTTGAGCTCGAGTCTACTCCAAAATCTCAGTAATAATACCT
APH1097	forward	GAAGAAAGATCTCGAGCTATGAGCGGTGAGATAAGTAGAG
APH1097	reverse	GAAGCTTGAGCTCGAGTCTATATTCTCATAGGCATAACGATGT
APH1098	forward	GAAGAAAGATCTCGAGCTATGGCTGTATTCGTGCATC
APH1098	reverse	GAAGCTTGAGCTCGAGTCTATGTCTGATGACTGAATATATCTACA
APH1099	forward	GAAGAAAGATCTCGAGCTATGCGCATACTGTTAATAGAAGAT
APH1099	reverse	GAAGCTTGAGCTCGAGTCTAGGCTTCTTCCGCATAATC
APH1100	forward	GAAGAAAGATCTCGAGCTATGGAAGAACACGGTGGTT
APH1100	reverse	GAAGCTTGAGCTCGAGTCTACCGTAGCGCCCTTG
APH1151	forward	GAAGAAAGATCTCGAGCTATGATTCTGTTTTTTGTTTCGTTAGCA
APH1151	reverse	GAAGCTTGAGCTCGAGTCTATCTCCAGTTAACACTCTCAC
APH1198	forward	GAAGAAAGATCTCGAGCTATGGGTGATGCTGTAGAAGTT
APH1198	reverse	GAAGCTTGAGCTCGAGTCTAAATTCCTAGAGCCAATCTGTT
APH1239	forward	GAAGAAAGATCTCGAGCTATGAGATCTAGAAGTAAGCTATTTTTAGGA
APH1239	reverse	GAAGCTTGAGCTCGAGTTTAAAGATGGTTTGTGTAATGAATTCCA
APH1263	forward	GAAGAAAGATCTCGAGCTATGGATAAAAAAGGTCCTAGAATAAACGA
APH1263	reverse	GAAGCTTGAGCTCGAGTTTAAGACCCACCACCAGTTT
APH1349	forward	GAAGAAAGATCTCGAGCTATGATTAGAGTGGGAATAAATGGC
APH1349	reverse	GAAGCTTGAGCTCGAGTTTATAAAAATTTCTGCGCTATTAGGG

Determination of T4SS substrates

A. phagocytophilum proteins that localized to the nucleus of HEK-293T and PLB-985 cells were tested for their ability to be secreted by the T4SS Dot/Icm system of *Coxiella burnetii* RSA439 avirulent phase II nine-mile strain using adenylate cyclase translocation assays.[27] Fusion proteins were created by cloning full-length coding regions or C-terminal 100 aa truncations to the *Bordetella pertussis* adenylate cyclase gene (*cyaA*). To achieve this, *C. burnetii* was transiently propagated in ACCM-2 axenic culture medium and transformed with the constructs (performed at the NIAID Rocky Mountain Laboratories [Hamilton, MT] by Paul Beare, Ph.D. and Charles L. Larson). Axenic *C. burnetii* was transformed by electroporation and cultured in ACCM-2 medium for 24 h followed by chloramphenicol selection.[28,29]The ability of the constructs to be secreted by the Dot/Icm system was determined by measuring changes in intracellular cAMP levels. CyaA fusion proteins that contain a T4SS signal are capable of being secreted and mediate a measurable increase in cAMP. *C. burnetii* transformants containing the *cyaA* constructs were used to infect THP-1 cells (a human myelomonocytic cell line) at an MOI of 100:1, and included *A. phagocytophilum* AnkA (APH_0740) and *Coxiella* vacuolar protein A (CvpA), both known T4SS substrates as positive controls. After 3 d, the cells were harvested, lysed and examined for cAMP production by enzyme immunoassay. Results were expressed as fold change in intracellular cAMP concentration compared to empty vector control (CyaA only) that lacked a T4SS signal; values >2 were considered positive for type 4 secretion; values between 1 and 2 were considered marginal.

Assay for detection of reactive oxygen species

Superoxide production was detected as described previously. [24] Briefly, HL-60 cells transfected by candidate *A. phagocytophilum* genes were incubated with 0.25 mM 2',7'-dichlorofluoresein diacetate (DCFH-DA) in PBS for 30 min at room temperature. 10^5 cells were stimulated in triplicate with 1 μg/mL phorbol 12-myristate 12-acetate (PMA) and fluorescence was measured every 2 min. The relative fluorescence units at 180 min were averaged and compared to unstimulated controls using a two-sided Student's *t*-test, α 0.05.

Results

In silico prediction of A. phagocytophilum proteins targeted to the host cell nucleus

Of 1,264 proteins and hypothetical proteins examined by the bioinformatics algorithm, 123 were identified by ProtComp as nuclear-localized; 3 of these were classified in PSORTb as potentially nuclear membrane-associated; after analysis of NLSdb and screening for NLSm and NLSb, 7 candidate proteins had a total Nuclear score >1 (Table 5.1). One candidate with a high ProtComp score for nuclear localization but that lacked a predicted NLS (APH_0805) was selected as a control. The known nuclear-translocated AnkA was not identified in this screen.

Table 5.1. Bioinformatic prediction of *A. phagocytophilum* nuclear-translocated proteins, by likelihood based on Final Score rank.

Locus Name	Acc. No.	Protein	Gene	Ranking	ProtComp	MultiLoc	NLS	NLS1	NLS2	Final Score
APH_0820[1]	YP_505397.1	hypothetical protein		10	2.1	0.94	1	1	0	2.44
APH_0847	YP_505424.1	hypothetical protein		22	2.1	0.97	0	1	1	1.97
APH_0382	YP_504988.1	HGE-14 protein		56	1.7	0.97	0	1	0	1.47
APH_0385	YP_504990.1	HGE-14 protein		75	2.1	0.94	0	1	0	1.44
APH_0455	YP_505057.1	HGE-14 protein		76	2.2	0.94	0	1	0	1.44
APH_0485	YP_505084.1	hypothetical protein		77	2.2	0.94	0	1	0	1.44
APH_0576	YP_505167.1	RNA polymerase sigma factor RpoD	*rpoD*	114	2.1	0.89	0	1	0	1.39
APH_0805[2]	*YP_505382.1*	*hypothetical protein*		*1891*	*2.1*	*0.96*	*0*	*0*	*0*	*0.96*

[1] not cloned

[2] selected as negative control

iTRAQ identification of A. phagocytophilum nuclear-translocated proteins

We detected 43 *A. phagocytophilum* proteins with an Aph:HL-60 ratio >1.2 in the nucleus of infected cells (Table 5.2), including the top hit, AnkA that is established to translocate into the nucleus. This approach allowed the identification of *A. phagocytophilum* proteins most likely to have been translocated into the nucleus and provided a more complete list of candidates to investigate out of the 1,264 *A. phagocytophilum* ORFs available for study. Of these 43 candidates, only AnkA was excluded from subsequent cloning and expression for *in vitro* nuclear localization studies.

Table 5.2. *Anaplasma phagocytophilum* proteins identified in the nuclear lysates of infected HL-60 cells by iTRAQ with ratios compared with uninfected cells of >1.2 and ranked by Unused ProtScore to identify high likelihood candidates for nuclear translocation.

Locus Name	Accession	Protein	Gene	ratios of labelled peptides[1]			Mean HL-60	Mean Aph	Ratio Aph:HL-60	Unused ProtScore
				115:114	116:114	117:114				
APH_0740	gi|88607707	ankyrin A	*ankA*	1.06	1.42	1.43	1.03	1.43	1.38	40.10
APH_1023	gi|88607105	DNA-directed RNA polymerase subunit beta; RNAP subunit beta	*rpoC*	1.04	1.50	1.39	1.02	1.45	1.42	32.10
APH_0240	gi|88606723	60 kDa chaperonin GroEL	*groEL*	1.00	1.88	1.93	1.00	1.90	1.90	30.60
APH_1024[2]	gi|88606872	DNA-directed RNA polymerase subunit beta; RNAP subunit beta	*rpoB*	1.02	1.33	1.27	1.01	1.30	1.28	23.40
APH_0906	gi|88606911	hypothetical protein APH_0906		1.05	1.25	1.22	1.02	1.24	1.21	20.70
APH_0278[2]	gi|88607578	translation elongation factor Tu; EF-Tu	*tufI*	1.18	1.93	1.83	1.09	1.88	1.73	20.20
APH_1099	gi|88607685	DNA-binding response regulator CtrA	*ctrA*	1.05	2.65	2.71	1.03	2.68	2.62	19.90
APH_0303	gi|88606699	DNA-directed RNA polymerase subunit alpha; RNAP subunit alpha	*rpoA*	1.16	1.91	1.84	1.08	1.88	1.74	15.90
APH_0784	gi|88606926	DNA-binding protein HU	*hup*	1.00	2.38	2.03	1.00	2.21	2.21	15.40
APH_0968	gi|88606840	ATP-dependent protease La	*lon*	1.01	1.62	1.44	1.01	1.53	1.52	15.10
APH_1100	gi|88606714	DNA-binding protein		0.99	2.82	2.44	0.99	2.63	2.64	13.80
APH_0469	gi|88607025	putative malonyl-CoA decarboxylase		1.04	1.27	1.28	1.02	1.27	1.24	12.60
APH_0445	gi|88607683	transcription elongation factor NusA	*nusA*	1.00	1.53	1.50	1.00	1.52	1.52	12.20
APH_0339	gi|88607311	putative thermostable metallocarboxypeptidase		1.11	1.59	1.58	1.05	1.58	1.50	9.60
APH_1239	gi|88607921	P44-15b outer membrane protein; major surface protein-2C	*p44-15b*	1.05	3.60	3.63	1.03	3.62	3.53	9.10

APH_0062	gi|88606901	hypothetical protein APH_0062		1.06	1.91	1.77	1.03	1.84	1.79	8.70
APH_1097	gi|88607712	DNA polymerase III, beta subunit	*dnaN*	1.14	1.33	1.30	1.07	1.32	1.23	6.60
APH_0135	gi|88606701	cold shock protein, CSD family		0.97	1.79	1.78	0.99	1.79	1.81	6.30
APH_0397	gi|88606909	30S ribosomal protein S2	*rpsB*	1.07	1.53	1.45	1.04	1.49	1.44	6.20
APH_1263	gi|88607227	translation initiation factor IF-3	*infC*	0.94	2.12	1.97	0.97	2.05	2.11	5.00
APH_0398	gi|88607503	Elongation factor Ts; EF-Ts	*tsf*	1.03	1.24	1.28	1.01	1.26	1.24	4.40
APH_1151	gi|88607101	hypothetical protein APH_1151		1.20	2.16	2.11	1.10	2.13	1.94	4.10
APH_0288	gi|88607038	50S ribosomal protein L29	*rpmC*	1.03	1.77	1.66	1.01	1.72	1.69	4.10
APH_1029	gi|88607731	transcription termination/antitermination factor NusG	*nusG*	0.94	1.65	1.72	0.97	1.69	1.74	4.00
APH_1027	gi|88607420	50S ribosomal protein L1	*rplA*	1.15	1.37	1.26	1.08	1.31	1.22	3.30
APH_0515	gi|88606905	expression regulator ApxR	*apxR*	0.99	2.02	1.94	0.99	1.98	2.00	3.20
APH_0097	gi|88606982	protein-export protein SecB	*secB*	1.11	1.81	1.67	1.06	1.74	1.64	3.00
APH_0292	gi|88606711	50S ribosomal protein L5	*rplE*	1.26	1.52	1.57	1.13	1.54	1.36	2.40
APH_0106	gi|88607568	riboflavin synthase, alpha subunit	*ribE*	0.98	2.04	2.00	0.99	2.02	2.04	2.30
APH_0280	gi|88607449	50S ribosomal protein L3	*rplC*	1.07	1.60	1.64	1.04	1.62	1.56	2.30
APH_0629	gi|88607793	malate dehydrogenase	*mdh*	0.98	1.25	1.21	0.99	1.23	1.24	2.30
APH_0160[2]	gi|88606875	putative thymidylate synthase, flavin-dependent, truncation		1.14	1.50	1.37	1.07	1.44	1.34	2.20
APH_0154	gi|88607134	serine hydroxymethyltransferase SHMT	*glyA*	1.06	1.29	1.26	1.03	1.27	1.24	2.10
APH_0971	gi|88607838	Trigger factor; TF	*tig*	1.04	1.37	1.28	1.02	1.32	1.30	2.00

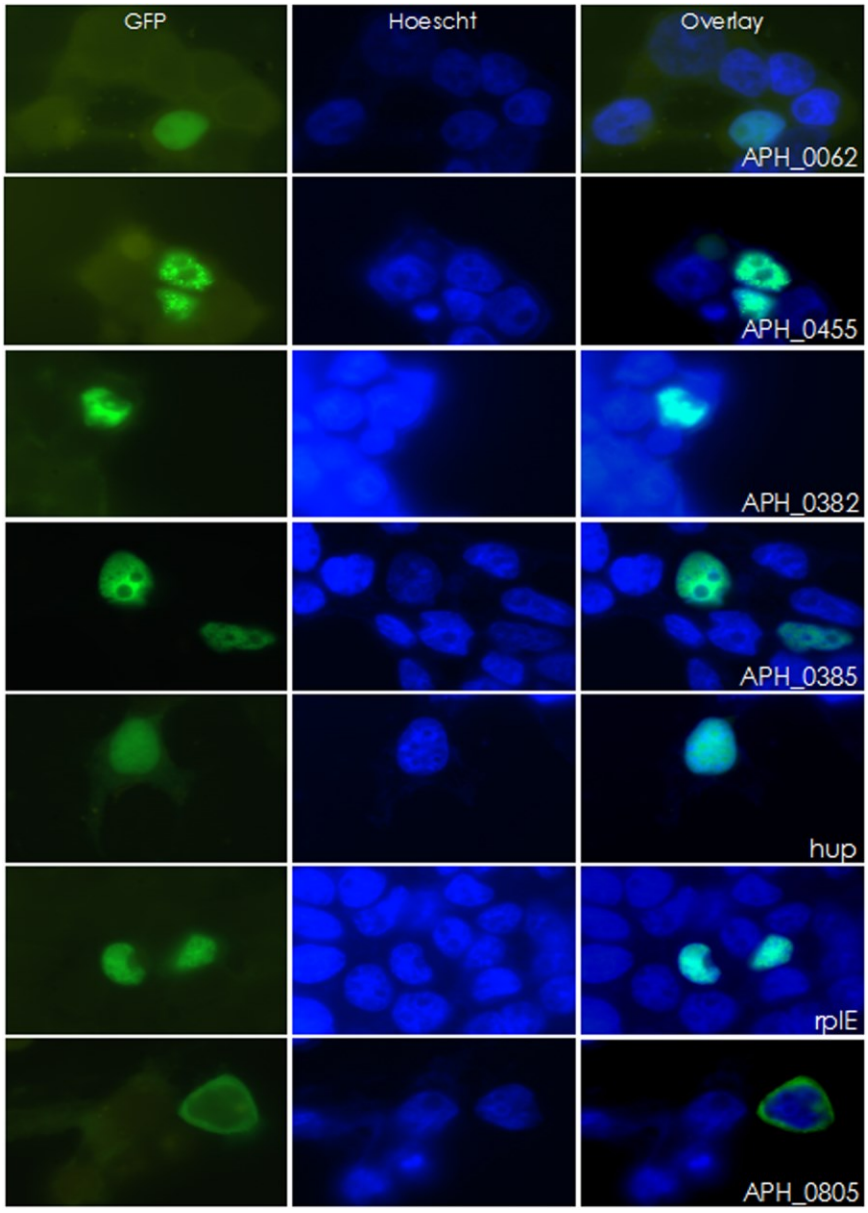
APH_0659	gi|88607183	antioxidant, AhpC/Tsa family		0.99	1.26	1.23	1.00	1.24	1.25	2.00
APH_1349	gi|88606948	glyceraldehyde-3-phosphate dehydrogenase, type I	*gap*	0.86	1.13	1.19	0.93	1.16	1.24	2.00
APH_1198	gi|88606994	2-oxoglutarate dehydrogenase, E2 component, dihydrolipoamide succinyltransferase	*sucB*	0.98	1.24	1.16	0.99	1.20	1.21	2.00
APH_1025	gi|88607605	50S ribosomal protein L7/L12	*rplL*	1.01	1.85	1.79	1.01	1.82	1.81	1.90
APH_0289[2]	gi|88607574	30S ribosomal protein S17	*rpsQ*	0.90	1.57	1.46	0.95	1.52	1.60	1.70
APH_1034[2]	gi|88607212	30S ribosomal protein S7	*rpsG*	1.16	1.42	1.41	1.08	1.41	1.31	1.50
APH_0196[2]	gi|88607673	Response Regulator NtrX, putative nitrogen assimilation regulatory protein	*ntrx*	0.93	1.70	1.50	0.97	1.60	1.66	1.40
APH_1333[2]	gi|88607617	transcription elongation factor GreA	*greA*	0.95	1.29	1.27	0.98	1.28	1.31	1.30
APH_1098	gi|88607131	3'-5' exonuclease family protein		1.12	1.30	1.29	1.06	1.30	1.22	1.30

[1] isobaric ion labels of nuclear lysates from: 114 and 115 – uninfected HL-60 cells; 116 and 117 – *A. phagocytophilum*-infected HL-60 cells.
[2] not cloned

In vitro nuclear localization

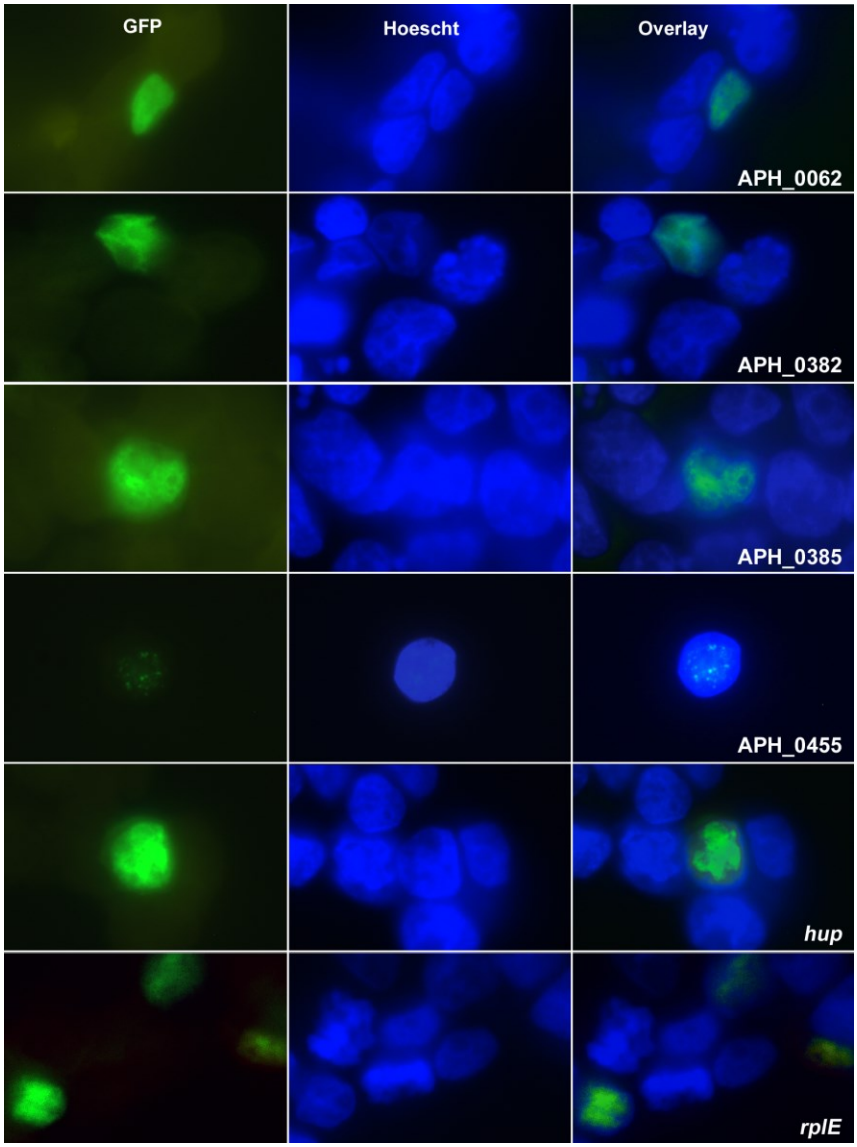
As an inclusive screen, and because contamination of nuclear preparations could not be excluded in iTRAQ studies, nuclear localization of proteins identified by bioinformatic methods or by iTRAQ mass spectrometry were confirmed by cloning the corresponding genes into a mammalian expression vector for expression as GFP fusion proteins. APH_0805 that was predicted to have nuclear localization yet lacked a predicted NLS and had a below-threshold Nuclear score was used as a non-translocating control. HEK-293T and PLB-985, a promyelocytic cell line, were transfected and examined for nuclear localization of the GFP-fusion proteins with DAPI nuclear counterstaining. Six of the 42 proteins tested (36 from iTRAQ profiling, 7 from the bioinformatic screen), translocated to the nucleus: APH_0062 (hypothetical protein), RplE (50S ribosomal protein L5 [APH_0292]), Hup (DNA-binding protein HU [APH_0783]), and APH_0455, APH_0382, and APH_0385 (all HGE-14) (Figure 5.1 and Supplemental Figure 5.2). APH_0278 (*tuf-1*; elongation factor Tu) was not cloned, but instead the identical APH_1032 (*tuf-2*; elongation factor Tu) was used but did not enter the nucleus. Nine proteins were either unable to be cloned or cloning was not attempted, including: APH_0160 (putative thymidylate synthase, flavin-dependent, truncation, partial); APH_0196 (nitrogen assimilation regulatory protein); APH_0289 (ribosomal protein S17 [*rpsQ*]); APH_0820 (hypothetical protein); APH_0906 (hypothetical protein); APH_1023 (DNA-directed RNA polymerase, beta subunit [*rpoC*]); APH_1024 (DNA-directed RNA polymerase, beta subunit [*rpoB*]); APH_1034 (ribosomal protein S7 [*rpsG*]) and APH_1333 (transcription elongation factor GreA).

Figure 5.1



Six *A. phagocytophilum* candidate genes were found to localize to the nucleus of HEK-293T cells. Candidate genes were fused to GFP and transfected into HEK-293T cells. 24 h post-transfection, cells were stained with DAPI and imaged. Of the 42 GFP-fusion proteins created, 6 localized to the nucleus. APH_0805 is shown here as an example of a protein that did not localize to the nucleus.

Supplemental Figure 5.2

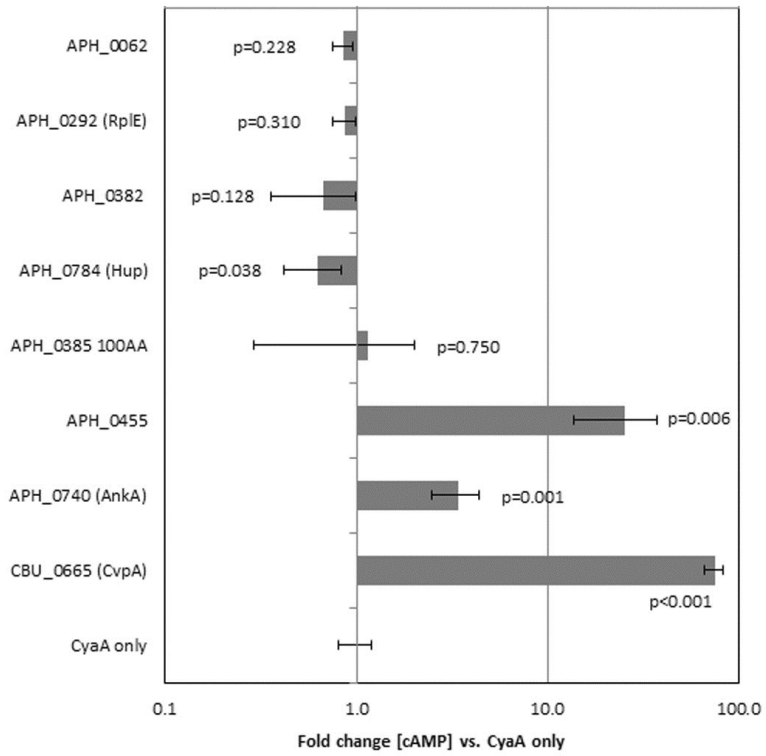


Nuclear translocation of six *A. phagocytophilum* proteins after nucleofection into PLB 985 promyelocytic cells. The six *A. phagocytophilum* candidate genes found to localize to the nucleus of HEK-293T cells were also nucleofected in PLB 985 promyelocytic leukemia cells after candidate genes were fused to GFP. 24 h post-transfection, cells were stained with DAPI and imaged. All 6 candidates also localized to the nucleus of PLB 985 granulocytes, reproducing similar morphologic distributions as in HEK 293T cells.

Determination of type 4 secretion substrates

Proteins identified to localize to the nucleus were further investigated to determine if they could be secreted by the T4SS of *Coxiella burnetii*, which is similar to that of *A. phagocytophilum*. T4SS substrate status was determined by the ability of the CyaA-fusion to exit *C. burnetii* and produce a measurable increase in cAMP concentrations with infection of THP-1 cells. Of the 6 genes tested only APH_0455 was identified to be a type 4 secretion substrate (Figure 5.2).

Figure 5.2

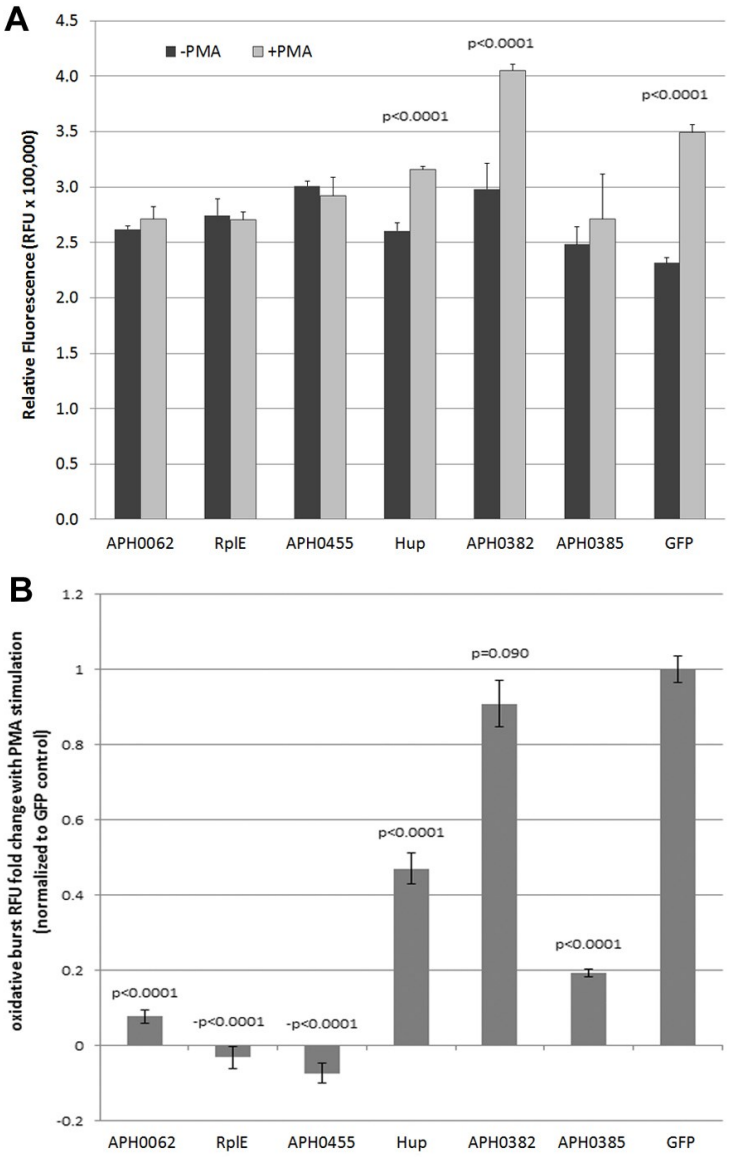


APH_0455 is secreted by the *Dot/Icm* T4SS of *C. burnetii*. Candidate genes were fused to *B. pertussis* adenylate cyclase (*cyaA*), transfected into *C. burnetii* and selected by chloramphenicol resistance. Transformed *C. burnetii* clones were then used to infect THP-1 cells. Three days post-transfection, THP-1 cells were assayed for cAMP production. Only those constructs that contain a T4SS signal sequence have measurable changes in cAMP production. The results represent the average of two separate experiments each with replicate tests. The p values were calculated based on comparisons with fold change of *C. burnetii* transformed by empty plasmid (CyaA only) using two-sided Student's t-tests, α=0.05. CBU_0655 is CvpA, a known T4SS substrate of *C. burnetii*.

Determination of oxidative burst after transfection and nuclear translocation

GFP-fusion constructs were transfected into HL-60 cells to determine their ability to alter the oxidative burst response. Unfortunately, the methods (electroporation, lipofectamine, lentivirus transduction) used to transfect the HL-60 cells (differentiated or undifferentiated), abrogated oxidative burst as compared with non-transfected cells. Thus, we compared results to PMA-stimulated oxidative burst in HL-60 cells transfected with the empty GFP plasmid as control. When RFU values of each unstimulated transfected control cell culture were compared to PMA-stimulated, significant oxidative burst, as seen with the GFP plasmid control, was observed only with Hup and APH_0382 (Figure 5.3A). When normalized to GFP plasmid transfection alone, APH_0062, RplE, APH_0455, Hup, and APH_0385 significantly repressed respiratory burst (Figure 5.3B). However, responses varied in intensity over several repeated experiments, likely in part due to the variable transfection efficiency obtained with HL-60 cells. These data suggest that one or more of these effectors could contribute to dampened production of reactive oxygen species.

Figure 5.3



Expression of putative nuclear effectors APH_0062, RplE and APH_0455 dampen PMA-stimulated reactive oxygen species production by HL-60 cells. HL-60 cells were transfected with 2 µg plasmid and assayed for respiratory burst 48 h later. A) The average of three replicates is displayed ± SEM at 180 min. B) The fold change was calculated by dividing the ratio of PMA stimulation of each transfectant to the GFP control. p values were calculated using Students t-tests.

Discussion

While considerable focus has been placed on AnkA as the primary nucleomodulin of *A. phagocytophilum*, it does not seem plausible that a single protein can account for the widespread transcriptional and phenotypic changes induced with infection. Using current bioinformatics tools and mass spectrometry, a number of other proteins encoded in the *A. phagocytophilum* genome were identified that could potentially localize to the host cell nucleus. To validate the candidate genes, GFP-fusion proteins were created and screened for nuclear localization within HEK-293T cells. This approach narrowed the list of target genes for further investigation to six. No candidate proteins were identified in both the bioinformatic screen and in the iTRAQ mass spectrometry analysis. If one assumes that the mass spectrometry data is accurate, the bioinformatic approach was ineffective at identifying features to predict nuclear localization for 3 of the six proteins shown capable of entering the nucleus; as a result, APH_0062 (cytoplasmic), *hup*, and *rplE* (both mitochondrial) were excluded from the bioinformatic identification because they were not assigned a nuclear localization. In contrast, no bioinformatic-predicted candidate appeared in the iTRAQ mass spectrometry analyses, suggesting limitations in sensitivity and/or contamination of nuclear preparations by non-nuclear localized proteins. Thus, the combination of both approaches increased the ability to identify and exclude candidates for further analysis. It is important to note that the screen will only identify those genes capable of entering the nucleus on their own accord via an identified or unidentified nuclear localization signal. Some proteins identified as present in the nucleus in the iTRAQ screen could indeed localize to the nucleus but might not be confirmed by transfection screens. A bacterial-derived protein shuttled into the nucleus as

a component of a protein complex, or one that possess an uncharacterized NLS, as is the case with AnkA, would not be identified. Furthermore, HEK-293T cells are not a model cell line for *A. phagocytophilum* infection and transfection of these proteins does not mimic infection, a much more complex process, therefore, confirmation of nuclear translocation in PLB-985 was performed.

Additionally, *A. phagocytophilum* is largely refractory to gene delivery by genetic transformation. Previous reports demonstrate *A. phagocytophilum* transformation using the Himar1 transposase system that introduces small GFP proteins or disrupts bacterial genes and consequently protein expression.[30,31] This process does not result in gene entry, but results in a library of mutant bacteria that can facilitate complex functional studies and insight into the importance of mutated genes for establishing or maintaining infection. However, directed mutation by homologous recombination has not yet been described for *A. phagocytophilum*.None-the-less, this relatively simple experiment yielded multiple candidate genes of interest for further investigation.

After narrowing the initial bioinformatic and iTRAQ list of candidate genes to six, we investigated the ability of these proteins to be secreted by the bacterium. For *A. phagocytophilum*, the most well characterized secretion mechanism is that of the T4SS. Because of this, we focused on whether or not these proteins could be secreted by a T4SS. As an obligate intracellular bacterium that resides solely in membrane-bound vacuoles of its host cells, *A. phagocytophilum*-secreted proteins very likely first enter the cytosol before translocation to the nucleus, but are unlikely to be detected outside of the host cell owing to the intracellular vacuolar membranes accessible to the bacterium. Thus, the *C. burnetii* Dot/Icm T4SS was used as a surrogate delivery system because *C.*

burnetii is capable of being transfected easily when cultivated in axenic medium but resides within host cell vacuoles when cultivated in mammalian cells. The *Dot/Icm* secretion system is compatible with that of *A. phagocytophilum* and, unless cultivated in specific axenic medium, *C. burnetii* is also an obligate intracellular bacterium residing within membrane-bound vacuoles. Using fusions with *B. pertussis* CyaA, one of six *A. phagocytophilum* candidate nuclear-localizing proteins was identified as a T4SS substrate. The remaining 5 did not appear to be secreted by the *Dot/Icm* system. Using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/), we determined the presence of putative Sec1 secretion signals in the genes encoding APH_0382, APH_0385, and APH_0455 (all HGE-14-like); experimental confirmation of this secretion mechanism was not further attempted.

Interestingly, APH_0382, APH_0385, and APH_0455 were shown to be differentially expressed between mammalian and tick cells. The transcription of each of these proteins was approximately 2.9 to 3.3-fold greater in HL-60 cells than ISE6 (tick) cells.[35] This suggests that these HGE-14-like proteins likely play a role in establishing or maintaining infection in mammalian cells. In fact, differential transcription of *A. phagocytophilum* genes plays a role in the life cycle of the bacterium in mammalian and tick cells.[35-38] APH_0784 (DNA binding protein HU), and APH_0292 (50S ribosomal protein L5) are among the 20 most abundant proteins expressed in infected *I. scapularis* salivary glands, [37] and both were found in nuclear lysates of infected HL-60 cells, yet predicted to localize to the mitochondrion and cytosol, respectively. We also identified the transcriptional regulator of p44/*msp2* genes, ApxR (APH_0515)[38] in nuclear lysates from *A. phagocytophilum* infected HL-60 cells, but at a low unused ProtScore. As ApxR

was unable to translocate to the nuclei of HEK293 cells, its presence indicates the potential for low level cytoplasmic contamination in the nuclear preparations. However, the overall level of cytoplasmic contamination is likely to be low since the most abundant *A. phagocytophilum* proteins in the P44/Msp2 family [35,37,38] were not abundant in nuclear lysates. Finally, APH_1235 is characterized as a specific marker of dense core infectious *A. phagocytophilum.*[36] It that is among the 20 most abundantly-expressed proteins in tick salivary glands,[37] is significantly upregulated in dense core cells with HL-60 cell infection,[36] and is believed to facilitate tick to mammal transmission. While predicted to localize to the nucleus by ProtComp v.6 and identified in infected HL-60 cell nuclear lysates, published data demonstrate the lack of nuclear localization. [36] Moreover, it lacked a recognized NLS and the iTRAQ unused score was low, suggesting low-level contamination from the host cytosol.

APH_0455 is of particular interest owing to its utilization of the T4SS to enter the cell and its translocation into the nucleus where it forms small aggregates and clusters dispersed unevenly throughout the nucleoplasm. APH-0455 is one of several HGE-14 proteins predicted to enter the nucleus, and APH-0455 has been described to have transmembrane domains that would predict it to be a type II membrane protein, and possesses 4 conserved 41 amino acid repeats followed by 2 similar truncated repeats.[32] This repeat region overlaps a region with a conserved Med15/ARC15 (pfam09606) domain. Med15/ARC105 domains are found as part of a family of sterol regulatory element binding proteins (SREBPs), transcription activators that regulate genes involved in cholesterol and fatty acid homeostasis. In humans, SREBPs bind CREB-binding protein (CBP)/p300 acetyltransferase that in turn affect chromatin structure and gene

transcription.[33] Whether APH_0455 plays a role in these critical pathways for *A. phagocytophilum* survival needs to be determined.[34]

Because of the candidate proteins' abilities to act as T4SS or Sec1 substrates and to localize to the nucleus, we sought to determine if they played a role in altering the phenotype of HL-60 cells, a commonly used cell model for *A. phagocytophilum* infection. Unfortunately, transfection of HL-60 cells with a variety of methods inconsistently altered oxidative burst capacity, and often the vehicle controls and transfection reagents were enough to abrogate responses. Despite the variable responses, we observed trends toward reduction of oxidative burst (Figure 5.3), but cannot currently conclude with certainty that these effectors play a role in limiting oxidative burst as shown for AnkA.[1]

For each of the *A. phagocytophilum* proteins that localized to the nucleus of HEK-293T and PLB-985 cells, it would be important confirm their presence in the nuclei of *A. phagocytophilum* infected cells visually or biochemically, and to potentially assess the effects of their absence in *A. phagocytophilum* among Himar1 transposase libraries.[35,36] Additionally, future studies will examine their role in transcriptional and functional changes in differentiated HL-60 cells, the preferred model for *A. phagocytophilum*-directed neutrophil reprogramming. Such studies will focus on transcriptional responses, functional assays and, given the role that AnkA plays during the course of infection, studies of nuclear protein-protein, DNA-protein, and RNA-protein interactions. The screening techniques modeled here using *A. phagocytophilum* will allow for a more focused approach to identify potential nucleomodulins and could facilitate studies of microbial nucleomodulin manipulation of host cell transcriptional programs.

These techniques are not limited to the *A. phagocytophilum* genome but can also be applied to other obligate intracellular bacteria. Using the same bioinformatics approaches (Supplemental Methods, supplemental Figure 5.1, Supplemental Tables 5.1 and 5.2), candidate genes were identified for other pathogens including, but not limited to: *Chlamydia trachomatis, Coxiella burnetii, Ehrlichia chaffeensis, Mycobacterium tuberculosis, Yersinia pestis, Legionella pneumophila, Francisella tularensis* and *Listeria monocytogenes.* Identification of new nucleomodulins in any one of these pathogens could add further insight as to how bacteria modulate their host cells and cause aberrant transcriptional reprogramming.

Conclusion

We used a combination of bioinformatic screens and iTRAQ *in vitro* identification of potential nuclear-translocated proteins to stratify and rapidly identify candidate nucleomodulins in *A. phagocytophilum,* an approach easily applied to other intracellular pathogens. By combining data gathered from bioinformatics prediction tools and iTRAQ, 50 *A. phagocytophilum* proteins were identified as potential nucleomodulins. Of the 50, we confirmed that six proteins were capable of localizing to the nucleus on their own, including APH_0455 that is also a T4SS substrate. The identification of novel nuclear translocated proteins provides additional support for the concept of nucleomodulin-mediated reprogramming of cellular functions that improve microbial fitness by promoting extended intracellular survival and more opportunities for transmission.

Conflict of Interest Statement. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements. Supported by grant R01AI044102 (JSD) from the US National Institutes of Allergy and Infectious Diseases, National Institutes of Health, and by subaward PO #SR00000759 (to JCGG) from Region III Mid-Atlantic Regional Center for Excellence grant U54 AI057168 from the NIAID (M.M. Levine, PI). The authors would like thank Paul Beare, Ph.D. and Charles Larson (Rocky Mountain Laboratories, NIAID, Hamilton, MT USA) for advice and performance of the T4SS assays; Kristen Rennoll-Bankert, Ph.D. (University of Maryland School of Medicine, Baltimore, MD USA) for technical assistance and support with cloning, expression and oxidative burst

assays; Carlos Borroto, M.S., and Wan Hsin (Cindy) Chen, M.S. (Johns Hopkins University) for assistance with cloning; and Robert Cole, Ph.D. (Johns Hopkins University School of Medicine) A. B. Mass Spectrometry/Proteomic Facility for help with iTRAQ experiments and analyses.

Supplemental Table 5.2. Bacterial strains and genomes selected for bioinformatic prediction of nuclear localized proteins.

Organism/strain	Group	Gram Staining	Size Mb	GC %	chr*	plasmids**	GenBank
Anaplasma phagocytophilum HZ	Alphaproteobacteria	-	1.47	42	1		CP000235
Brucella abortus biovar 1 str. 9-941	Alphaproteobacteria	-	3.29	57	2		AE017223
Chlamydia trachomatis D/UW-3/CX	Chlamydiales	-	1.05	41	1	1	CP000051
Chlamydophila pneumoniae AR39	Chlamydiales	-	1.23	40	1		AE001273
Coxiella burnetii RSA 493	Gammaproteobacteria	-	2.03	43	1	1	AE016828
Ehrlichia chaffeensis str. Arkansas	Alphaproteobacteria	-	1.18	30	1		CP000236
Francisella tularensis subsp. holarctica	Gammaproteobacteria	-	1.90	32	1		AM233362
Legionella pneumophila subsp.	Gammaproteobacteria	-	3.40	38	1		AE017354

Listeria monocytogenes str. 4b F2365	Firmicutes	+	2.91	37	1		AE017262
Mycobacterium tuberculosis CDC1551	Actinobacteria	+	4.40	66	1		AE000516
Rickettsia prowazekii str. Madrid E	Alphaproteobacteria	-	1.11	29	1		AJ235269
Yersinia pestis CO92	Gammaproteobacteria	-	4.83	48	1	3	AL590842

*number of chromosomes; ** number of plasmids

Supplemental Table 5.3. Prediction of bacterial proteins for nuclear localization and the presence of an nuclear localization sequence (NLS).

Organism/strain	DNA	RefSeq	ORFs	Nuclear ProtComp	NLS MultiLoc
Anaplasma phagocytophilum HZ	Chromosome	NC_007797	1264	120	4
Brucella abortus biovar 1 str. 9-941	Chromosome I	NC_006932	2030	99	5
	Chromosome II	NC_006933	1055	47	6
Chlamydia trachomatis D/UW-3/CX	Chromosome	NC_000117	895	58	7
	Plasmid pCTA	NC_007430	8	1	0
Chlamydophila pneumoniae AR39	Chromosome	NC_002179	1112	74	10
Coxiella burnetii RSA 493	Chromosome	NC_002971	2016	174	29
	Plasmid pQpH1	NC_004704	36	5	0
Ehrlichia chaffeensis str. Arkansas	Chromosome	NC_007799	1105	141	12
Francisella tularensis subsp. holarctica	Chromosome	NC_007880	1754	92	9
Legionella pneumophila subsp. pneumophila	Chromosome	NC_002942	2942	163	29
Listeria monocytogenes str. 4b F2365	Chromosome	NC_002973	2821	97	16
Mycobacterium tuberculosis CDC1551	Chromosome	NC_002755	4189	247	14
Rickettsia prowazekii str. Madrid E	Chromosome	NC_000963	835	70	11
Yersinia pestis CO92	Chromosome	NC_003143	3885	173	6
	Plasmid pCD1	NC_003131	72	7	0
	Plasmid pPCP1	NC_003132	9	2	0
	Plasmid pMT1	NC_003134	101	9	0
TOTALS			**26129**	**1579**	**158**

References:

1. Banerjee R, Anguita J, Roos D, Fikrig E. Cutting edge: Infection by the agent of human granulocytic ehrlichiosis prevents the respiratory burst by down-regulating *gp91phox*. *J Immunol*. 2000;164(8):3946-3949.

2. Borjesson DL, Kobayashi SD, Whitney AR, Voyich JM, Argue CM, Deleo FR. Insights into pathogen immune evasion mechanisms: *Anaplasma phagocytophilum* fails to induce an apoptosis differentiation program in human neutrophils. *J Immunol*. 2005;174(10):6364-6372.

3. Carlyon JA, Fikrig E. Mechanisms of evasion of neutrophil killing by *Anaplasma phagocytophilum*. *Curr Opin Hematol*. 2006;13(1):28-33.

4. Carlyon JA, Chan WT, Galan J, Roos D, Fikrig E. Repression of *rac2* mRNA expression by *Anaplasma phagocytophila* is essential to the inhibition of superoxide production and bacterial proliferation. *J Immunol*. 2002;169(12):7009-7018.

5. Choi KS, Park JT, Dumler JS. *Anaplasma phagocytophilum* delay of neutrophil apoptosis through the p38 mitogen-activated protein kinase signal pathway. *Infect Immun*. 2005;73(12):8209-8218.

6. Sinclair SH, Rennoll-Bankert KE, Dumler JS. Effector bottleneck: Microbial reprogramming of parasitized host cell transcription by epigenetic remodeling of chromatin structure. *Front Genet*. 2014;5:274.

7. Caturegli P, Asanovich KM, Walls JJ, et al. ankA: An *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infect Immun*. 2000;68(9):5277-5283.

8. Park J, Kim KJ, Choi KS, Grab DJ, Dumler JS. *Anaplasma phagocytophilum* AnkA binds to granulocyte DNA and nuclear proteins. *Cell Microbiol*. 2004;6(8):743-751.

9. Garcia-Garcia JC, Rennoll-Bankert KE, Pelly S, Milstone AM, Dumler JS. Silencing of host cell *CYBB* gene expression by the nuclear effector AnkA of the intracellular pathogen *Anaplasma phagocytophilum*. *Infect Immun*. 2009;77(6):2385-2391.

10. Garcia-Garcia JC, Barat NC, Trembley SJ, Dumler JS. Epigenetic silencing of host cell defense genes enhances intracellular survival of the rickettsial pathogen *Anaplasma phagocytophilum*. *PLoS Pathog*. 2009;5(6):e1000488.

11. Rennoll-Bankert KE, Dumler JS. Lessons from A*naplasma phagocytophilum*: Chromatin remodeling by bacterial effectors. *Infect Disord Drug Targets*. 2012;12(5):380-387.

12. Dunning Hotopp JC, Lin M, Madupu R, et al. Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet*. 2006;2(2):e21.

13. Rikihisa Y, Lin M, Niu H. Type IV secretion in the obligatory intracellular bacterium *Anaplasma phagocytophilum*. *Cell Microbiol*. 2010;12(9):1213-1221.

14. Lin M, den Dulk-Ras A, Hooykaas PJ, Rikihisa Y. *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cell Microbiol*. 2007;9(11):2644-2657.

15. Al-Khedery B, Lundgren AM, Stuen S, et al. Structure of the type IV secretion system in different strains of *Anaplasma phagocytophilum*. *BMC Genomics*. 2012;13:678-2164-13-678.

16. IJdo JW, Carlson AC, Kennedy EL. *Anaplasma phagocytophilum* AnkA is tyrosine-phosphorylated at EPIYA motifs and recruits SHP-1 during early infection. *Cell Microbiol*. 2007;9(5):1284-1296.

17. Ohashi N, Zhi N, Lin Q, Rikihisa Y. Characterization and transcriptional analysis of gene clusters for a type IV secretion machinery in human granulocytic and monocytic ehrlichiosis agents. *Infect Immun*. 2002;70(4):2128-2138.

18. Donnes P, Hoglund A. Predicting protein subcellular localization: Past, present, and future. *Genomics Proteomics Bioinformatics*. 2004;2(4):209-215.

19. Gardy JL, Laird MR, Chen F, et al. PSORTb v.2.0: Expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*. 2005;21(5):617-623.

20. Hicks GR, Raikhel NV. Protein import into the nucleus: An integrated view. *Annu Rev Cell Dev Biol*. 1995;11:155-188.

21. Hoglund A, Donnes P, Blum T, Adolph HW, Kohlbacher O. MultiLoc: Prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics*. 2006;22(10):1158-1165.

22. Nair R, Carter P, Rost B. NLSdb: Database of nuclear localization signals. *Nucleic Acids Res*. 2003;31(1):397-399.

23. Chen F, Mackey AJ, Stoeckert CJ,Jr, Roos DS. OrthoMCL-DB: Querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res*. 2006;34(Database issue):D363-8.

24. Rennoll-Bankert KE, Sinclair SH, Lichay MA, Dumler JS. Comparison and characterization of granulocyte cell models for *Anaplasma phagocytophilum* infection. *Pathog Dis*. 2013.

25. Ellison MA, Thurman GW, Ambruso DR. Phox activity of differentiated PLB-985 cells is enhanced, in an agonist specific manner, by the PLA2 activity of Prdx6-PLA2. *Eur J Immunol*. 2012;42(6):1609-1617.

26. Pedruzzi E, Fay M, Elbim C, Gaudry M, Gougerot-Pocidalo MA. Differentiation of PLB-985 myeloid cells into mature neutrophils, shown by degranulation of terminally differentiated compartments in response to N-formyl peptide and priming of superoxide anion production by granulocyte-macrophage colony-stimulating factor. *Br J Haematol*. 2002;117(3):719-726.

27. Larson CL, Beare PA, Howe D, Heinzen RA. *Coxiella burnetii* effector protein subverts clathrin-mediated vesicular trafficking for pathogen vacuole biogenesis. *Proc Natl Acad Sci U S A*. 2013;110(49):E4770-9.

28. Beare PA, Howe D, Cockrell DC, Omsland A, Hansen B, Heinzen RA. Characterization of a *Coxiella burnetii ftsZ* mutant generated by Himar1 transposon mutagenesis. *J Bacteriol*. 2009;191(5):1369-1381.

29. Lodes MJ, Mohamath R, Reynolds LD, et al. Serodiagnosis of human granulocytic ehrlichiosis by using novel combinations of immunoreactive recombinant proteins. *J Clin Microbiol*. 2001;39(7):2466-2476.

33. Yang F, Vought BW, Satterlee JS, et al. An ARC/mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature*. 2006;442(7103):700-704.

34. Lin M, Rikihisa Y. *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* lack genes for lipid A biosynthesis and incorporate cholesterol for their survival. *Infect Immun*. 2003;71(9):5324-5331.

35. Nelson CM, Herron MJ, Felsheim RF, et al. Whole genome transcription profiling of *Anaplasma phagocytophilum* in human and tick host cells by tiling array analysis. *BMC Genomics*. 2008;9:364-2164-9-364.

36. Troese MJ, Kahlon A, Ragland SA, et al. Proteomic analysis of *Anaplasma phagocytophilum* during infection of human myeloid cells identifies a protein that is pronouncedly upregulated on the infectious dense-cored cell. *Infect Immun*. 2011;79(11):4696-4707.

37. Mastronunzio JE, Kurscheid S, Fikrig E. Postgenomic analyses reveal development of infectious *Anaplasma phagocytophilum* during transmission from ticks to mice. *J Bacteriol*. 2012;194(9):2238-2247.

38. Wang X, Cheng Z, Zhang C, Kikuchi T, Rikihisa Y. *Anaplasma phagocytophilum* p44 mRNA expression is differentially regulated in mammalian and tick host cells: Involvement of the DNA binding protein ApxR. *J Bacteriol*. 2007;189(23):8651-8659.

Chapter 6: Comparison and characterization of granulocyte cell models for *Anaplasma phagocytophilum* infection

Rennoll-Bankert KE, Sinclair SH, Lichay MA, Dumler JS. (2014). Comparison and characterization of granulocyte cell models for *Anaplasma phagocytophilum* infection. Pathog Dis. Jun;71(1):55-64 doi: 10.1111/2049-632X.12111

Abstract

Anaplasma phagocytophilum, an obligate intracellular bacterium, modifies functions of its *in vivo* host, the neutrophil. The challenges of using neutrophils *ex vivo* necessitate cell line models. However, cell line infections do not currently mimic *ex vivo* neutrophil infection characteristics. To understand these discrepancies, we compared infection of cell lines to *ex vivo* human neutrophils and differentiated hematopoietic stem cells with regard to infection capacity, oxidative burst, host defense gene expression, and differentiation. Using established methods, marked *ex vivo* neutrophil infection heterogeneity was observed at 24-48 h necessitating cell sorting to obtain homogeneously-infected cells at levels observed *in vivo*. Moreover, gene expression of infected cell lines differed markedly from the prior standard of unsorted infected neutrophils. Differentiated HL-60 cells sustained similar infection levels to neutrophils *in vivo* and closely mimicked functional and transcriptional changes of sorted infected neutrophils. Thus, care must be exercised using *ex vivo* neutrophils for *A. phagocytophilum* infection studies since a major determinant of transcriptional and functional changes among all cells was the intracellular bacteria quantity. Furthermore, comparisons of *ex vivo* neutrophils and the surrogate HL-60 cell model allowed the determination that specific cellular functions and transcriptional programs are targeted by the bacterium without significantly modifying differentiation.

Introduction

The obligate intracellular pathogen, *Anaplasma phagocytophilum,* survives and propagates primarily within neutrophils by reprogramming critical granulocyte functions. This reprogramming includes delayed neutrophil apoptosis that allows time for bacterial replication, [1-3] increased recruitment and clustering of neutrophils which promotes bacterial dissemination and inflammatory response, [4-6] and impaired host defenses, such as reduced NADPH oxidase superoxide anion production that permits intracellular survival. [7-11] These modifications occur with active intracellular replication and with changes in host gene transcription. For example, reduced NADPH oxidase activation is in part attributed to decreased granulocyte *CYBB* transcription. [7,12] The *A. phagocytophilum* nucleomodulin AnkA, binds to the *CYBB* promoter and downregulates its expression. [13] *A. phagocytophilum* infection also leads to downregulation of host granulocyte defense genes including catalase (*CAT*), cathepsin G (*CTSG*), defensin alpha 4 (*DEFA4*), elastase 2 (*ELA2*), myeloperoxidase (*MPO*), proteinase 3 (*PRTN3*), and bactericidal/permeability-increasing protein (*BPI*), among others. [13,14] Moreover, delayed apoptosis is in part maintained by upregulated transcription of *BCL2* family genes, whereas neutrophil recruitment is enhanced by upregulated chemokine gene transcription, especially *IL8.* [15-19] Complex and coordinated functional changes such as reduced adhesion of infected neutrophils to endothelial cells, their transmigration through endothelium, enhanced degranulation, and impaired phagocytosis are phenotypic expressions that resemble neutrophil progenitors more than terminally differentiated neutrophils. [10,20,21] Yet, the coordinated subversion of each function provides a significant fitness advantage for intracellular survival in neutrophils and subsequent acquisition by tick blood meal.

Understanding the genome-wide basis for transcriptional and epigenetic subversion of complex phenotypic functions by *A. phagocytophilum* will require infections in *ex vivo* neutrophils or other adequate tractable surrogate cell models.

Investigation of functional alterations owing to *A. phagocytophilum* infection is most relevant in the natural mammalian target cell, the neutrophil. However, neutrophils present difficult challenges for experimental studies *ex vivo*, specifically with regard to pre-programmed apoptosis and short *ex vivo* life span, inability to manipulate transcription, and difficulty with transfection for expression of exogenous proteins or silencing of gene expression. As a result, investigation of the functional effects of *A. phagocytophilum* infection is most often conducted in granulocyte cell line models including HL-60, THP-1, and NB4 cells. [13,14,18] Although cell lines have substantially contributed to studies of the functional effects of *A. phagocytophilum* infection, each cell model has deficits for study of neutrophil differentiation or function. Moreover, *ex vivo* neutrophil transcriptional responses with *A. phagocytophilum* infection do not yield the same results as observed in granulocyte cell lines. [15,16,18,19,22] No study has examined why such discrepancies exist or which cell line(s) most closely mimic responses and behavior of infected neutrophils. Additionally, the PLB-985 human myelomonoblastic cell line and differentiated human hematopoietic stem cells (HSCs) have yet to be investigated as *in vitro* models of infection. HSCs hold promise for modeling *A. phagocytophilum* infection because they are primary cells that lack neoplastic mutations and can be transfected to express exogenous proteins or silence endogenous gene expression. To determine which cell line models could accurately be used to study genome-wide modulation of transcriptional programs that belie neutrophil reprogramming with *A. phagocytophilum*

infection, we sought to test i) why neutrophil transcriptional responses differ from those in granulocyte cell models, and ii) how accurately these cell line models reflect the differentiation, functional, and transcriptional effects of infection in neutrophils.

Materials and Methods

Cell lines and cell culture

Human neutrophils were obtained from EDTA anti-coagulated peripheral blood under a protocol approved by the Johns Hopkins Medical Institutions IRB. Dextran-sedimented, leukocyte-rich plasma was centrifuged through Ficoll-Paque density gradients, mononuclear cells were discarded, and residual erythrocytes were lysed in hypotonic buffer to obtain purified neutrophils. The promyelocytic HL-60 (ATCC CCL-240) and acute monoblastic/myelomonocytic leukemia THP-1 (ATCC TIB-202) cell lines were purchased from American Type Culture Collection (Manassas, VA). The acute promyelocytic leukemia NB4 and myelomonoblastic leukemia PLB-985 cell lines were gifts from Dr. Alan Friedman (The Johns Hopkins University, Baltimore, MD, USA) and Dr. Frank DeLeo (Rocky Mountain Laboratories, NIAID, Hamilton, MT, USA), respectively. The neutrophils were maintained and cell lines were grown in RPMI 1640 medium (Hyclone, Thermo Fisher Scientific, Waltham, MA) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, Waltham, MA) and Glutamax (Life Technologies, Carlsbad, CA). HL-60, NB4, and THP-1 cells were differentiated 5 days prior to infection with 1 μM all-*trans* retinoic acid (ATRA) [23-25] while PLB-985 cells were differentiated with 1.25% dimethylsulfoxide (DMSO). [26] HSCs purchased from Promocell GmbH (Heidelberg, Germany) were expanded in and differentiated for 14 days prior to infection in hematopoietic progenitor cell expansion medium DXF supplemented with cytokine E mix (Promocell GmbH, Heidelberg, Germany). All cells were grown in a humidified incubator at 37°C with 5% CO_2. Cell density was kept $<10^6$ cells/mL by diluting with fresh medium.

A. phagocytophilum isolation and infection

High passage *A. phagocytophilum* Webster strain was re-activated for virulence by equine passage and re-isolated in HL-60 cells. [27] Briefly, two *A. phagocytophilum* seronegative and PCR-negative horses were inoculated with 10^6 cell-free bacteria. Eight days post inoculation both horses were febrile (103°F and 104°F) and blood containing approximately 10^6 infected cells was used to inoculate a second set of horses. By day 7 the horses were febrile and blood was obtained for isolation in HL-60 cell culture. These experiments were conducted under an IACUC-approved protocol at the University of California, Davis.

The re-activated *A. phagocytophilum* was maintained in HL-60 cells as previously described. [28] For all experiments *A. phagocytophilum* was passed less than 10 times *in vitro*. The GFP- and mCherry-transgenic *A. phagocytophilum* HGE1 strains (a gift from Dr. Ulrike Munderloh, University of Minnesota, St. Paul, MN, USA) were maintained in HL-60 cells supplemented with 100 μg/mL spectinomycin. [29] To obtain cell-free *A. phagocytophilum,* 90-95% infected HL-60 cells were centrifuged at 500xg for 5 min. Cells were suspended in 1X PBS and sonicated two times using output 4 of a Branson Sonifier 250 (Branson Ultrasonics, Danbury, CT) for 15 sec. Cell debris was removed by centrifugation at 1,000xg for 10 min. The bacteria-containing supernatant was then centrifuged at 13,000xg for 30 min and bacteria were suspended in RPMI 1640 medium.

A multiplicity of infection (MOI) of 25 bacteria per cell was used for infection following methods modified from Borjesson et al. and Goodman et al. [15,17] Infection levels were determined by microscopic examination of Wright-Giemsa-stained (Protocol HEMA3, Thermo Fisher Scientific, Waltham, MA) cytocentrifuged cells and by real time

quantitative PCR (qPCR) at 72 h post-infection for cell lines and HSCs, and 24 h post-infection for neutrophils. qPCR was performed with DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA) extracted DNA and an *msp2*/*ACTB* 5' nuclease assay. [6] A standard curve based on cloned *msp2* was generated and used to calculate the number of bacteria per cell. There are >100 copies of *msp2* per *A. phagocytophilum* genome.

Fluorescence activated cell sorting

Prior transcriptional studies in *A. phagocytophilum*-infected neutrophils used intervals up to 24 h during which bacterial replication (12-24 h doubling time) was likely to be minimal. [16,17,19] Despite using conditions applied by others to study differential neutrophil transcription with *A. phagocytophilum* infection, [15,17] morphologic and qPCR methods showed a highly heterogeneous population comprised of small numbers of infected cells and large numbers of apparently uninfected cells. To preclude the confounding variable introduced by using heterogeneously-infected neutrophil cultures, after 24 h infection, 10^7 neutrophils infected with GFP or mCherry transgenic *A. phagocytophilum* were washed in RPMI 1640 medium containing 1% FBS, examined by flow cytometry and a homogeneously-infected population was isolated by fluorescence activated cell sorting (FACS) into cell lysis buffer (RNeasy Mini Prep, Qiagen, Valencia, CA) using a MoFlo sorter (Beckman Coulter, Brea, CA). RNA extraction, cDNA synthesis RT-PCR and qPCR were performed as described below. Non-viable cells were excluded by gating and among a population of 5 x 10^6 viable neutrophils, 10^4 (0.2%) of the most fluorescent cells were sorted for analysis of transcription (Fig. S6.2). As a control, unsorted infected neutrophil cultures were also examined.

RNA isolation, Reverse Transcriptase-PCR, and quantitative PCR

RNA was extracted using RNeasy RNA Extraction Kit (Qiagen, Valencia, CA) and cDNA was synthesized using iScript cDNA synthesis kit (Bio-Rad, Hercules, CA) according to manufacturer's protocols. Reverse transcriptase PCR (RT-PCR) was performed with VeriQuest SYBR Green qPCR Master Mix (Affymetrix, Santa Clara, CA) in a CFX384 Multicycler (Bio-Rad, Hercules, CA) (Table S6.1). The thermal cycling conditions included 95°C for 3 min followed by 40 cycles of amplification at 95°C for 10 sec and 55°C for 30 sec. A melt curve analysis was performed to confirm amplification of a single product for each primer pair. A panel of 12 reference genes (Table S6.1) was tested for each cell line to determine which were stably expressed (geNorm M score ≤0.5, Qbase PLUS, Biogazelle, Zwijnaarde, Belgium) with infection and differentiation. Fold change expression of defense genes was calculated using the ΔΔCt method [30] normalized to selected reference genes and relative to uninfected cells.

Because FACS separated more heavily infected neutrophils directly into RNA cell lysis buffer, we tested DNA and RNA from 24 h *A. phagocytophilum*-infected HL-60 cells by qPCR and RT-qPCR in the *msp2/ACTB* 5' nuclease assay [6] to determine the ratio of *msp2* transcripts to *msp2* genes per bacterium. This ratio was then used to quantify the number of bacteria per neutrophil based on *msp2* transcript quantitation before and after flow cytometric sorting of infected neutrophils.

Oxidative burst

Production of superoxide anion was measured by incubating 10^6 cells with 0.25 mM 2′,7′-dichlorohydrofluorescein diacetate (DCFH-DA) in PBS for 30 min at room temperature. 10^5 cells were then either stimulated with 1 μg/mL phorbol 12-myristate 13-acetate (PMA) or not stimulated and fluorescence was measured every 2 min with a

Victor^3 1420 Multilabel Counter (PerkinElmer, Norwalk, CT). The relative light units at 2 h in three independent experiments were averaged and compared using a two-sided Student's t-test, α 0.05.

Immunophenotyping of myeloid differentiation

Expression of cell surface markers was used to characterize the differentiation of cell lines when exposed to ATRA, DMSO, and *A. phagocytophilum*. 10^5 cells were incubated in 2% FBS in PBS with fluorescently-tagged antibodies (Table S6.2) for 20 min and then fixed with BD FACS Lysing solution (Becton Dickinson, Franklin Lakes, NJ) for 10 min. Cells were washed and resuspended in 2% FBS in PBS. 10,000 events were counted on a BD Fortessa flow cytometer (Becton Dickinson, Franklin Lakes, NJ). Cells were first gated and separated by viability as determined by Live/Dead staining (Invitrogen, Grand Island, NY). The fluorescence minus one (FMO) method was used to determine standard gates for each marker. Expression of surface markers CD34 (hematopoietic progenitor cell antigen, expressed on myeloblasts through promyelocytes), CD66b (CEACAM8, expressed on promyelocytes through neutrophils), CD11b (ITGAM, expressed on myelocytes through neutrophils), CD16 (FCGR3, expressed on metamyelocytes through neutrophils), CD35 (complement receptor [CR] 1, expressed on band neutrophils through neutrophils), and CD10 (neprilysin, expressed on mature neutrophils) were examined to determine the degree of differentiation (Fig. 6.4A). CD162 (selectin P ligand) was also investigated because of its role as attachment ligand and in invasion of *A. phagocytophilum*.[28,31]

The geometric mean of three independent experiments was calculated and averaged for each marker per cell type, differentiation status, and time point. Mean

fluorescence intensity was normalized across each cell type for each surface marker to allow comparisons within cell lines. For example, the geometric mean fluorescence of CD34 for all PMN samples was summed, and the proportion of CD34 expression for each condition (differentiated, not differentiated, infected, uninfected at 3, 24, and 48 h) was calculated based on the total sum fluorescence of that marker in PMNs. This was repeated for each marker within each cell line. The proportion of expression for each marker was then weighted based on the specific differentiation marker, such that those markers reflecting a less differentiated phenotype received a negative weight (-1 for CD34) or no weight (+1 for CD66b), and markers associated with an increasingly neutrophil-like differentiation were weighted more heavily (+2 for CD11b, +3 for CD16, +4 for CD35, and +5 for CD10). The weighted scores were summed for each cell type to score the relative degree of differentiation across the variables (differentiated, not differentiated, infected, uninfected, and time post infection if applicable). Since the final scores were not normally distributed, they were ranked for nonparametric analysis. The most differentiated cells received the highest rank (22 out of 22) and the least differentiated the lowest rank (1 out of 22). CD162 surface expression was analyzed separately from the differentiation markers. Similarly, the proportion of CD162 expression for each condition was ranked and cells with highest proportion received the highest rank. Ranked scores were compared using a two-sided Mann-Whitney test, with α=0.05.

Results

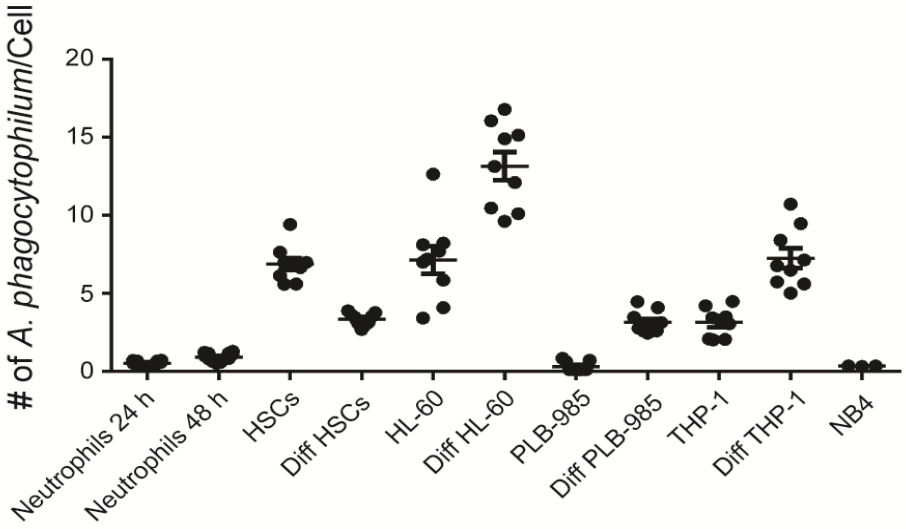
A. phagocytophilum infection level differs among granulocyte models

Peripheral blood neutrophils infected by *A. phagocytophilum in vivo*, such as in humans, horses, or mouse models, generally develop single morulae containing approximately 10 bacterial profiles.[32] Thus, the degree to which each cell line became infected with *A. phagocytophilum* was first determined. *A. phagocytophilum* re-activated for virulence by passage in horses was used for *in vitro* infection to preclude the effects of bacterial adaption to HL-60 cells with prolonged passage. Neutrophils as well as undifferentiated and differentiated cell lines were infected with an MOI of 25 bacteria per cell. At 24 and 48 h post infection for neutrophils and 72 h post infection for cell lines, cells were stained with Wright-Giemsa and qPCR was performed on DNA from 10^5 cells. Neutrophil infection was assessed at 3, 24 and 48 h because the cells became apoptotic after 48 h post infection; uninfected neutrophils were apoptotic at 24 h. Cell lines were assessed at 72 h to obtain homogeneous infections. Despite using methods for infection as described in other neutrophil transcription studies, we found that by 48 h, neutrophils were infected with an average of less than 1 bacterium per cell (Fig. 6.1) as detected by qPCR. Wright-Giemsa stains showed a heterogeneous mixture that included only a very small number of infected but viable cells and a large proportion of uninfected apoptotic neutrophils (Fig. S6.1A), in line with our and other previous reports.[2,3,33-35] After sorting the most fluorescent 0.2% of cells, the mean was 40.4 (± 6.0 s.e.m.) *A. phagocytophilum* per neutrophil.

All cell lines differentiated with ATRA or DMSO and HSCs differentiated as per protocol developed morphology that was more neutrophil-like with multi-lobed nuclei and increased cytoplasmic granules as compared to undifferentiated cells (Fig. S6.1B).

Differentiated HL-60 cells had the highest infection burden with approximately 12 bacteria per cell (Fig. 6.1 and S6.1B), similar to that anticipated with infection *in vivo*. Undifferentiated PLB-985 and NB4 cells had the lowest number of bacteria per cell, and in some cases, infection was undetectable at 72 h (Fig. 6.1 and S6.1B). Because infection could not be consistently detected in NB4 cells, analysis of this cell line was not pursued.

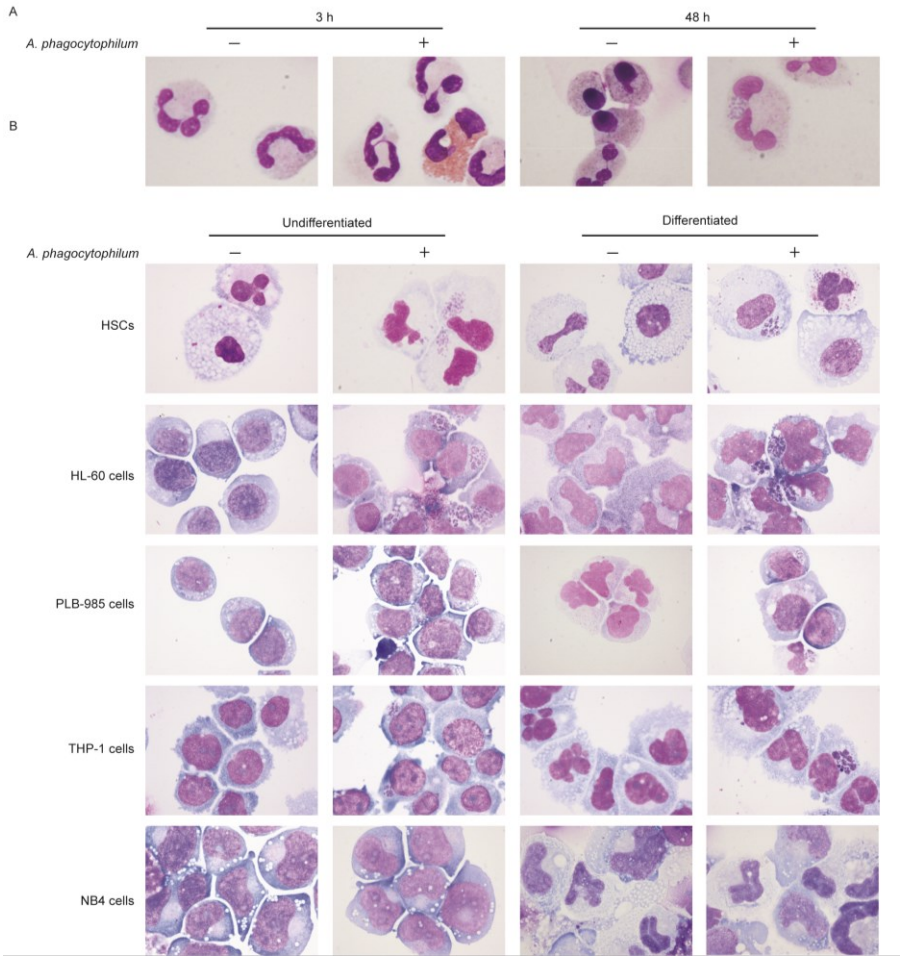
Figure 6.1



***A. phagocytophilum* infects granulocytes at differing efficiencies.**

DNA was extracted from cells at 24 and 48 h for PMNs and 72 h for other cells. The DNA was analyzed by qPCR for the presence of *A. phagocytophilum* infection with *msp2/B-actin* 5' nuclease assay and the number of *A. phagocytophilum* per cell was determined. Infection was performed in quadruplicate for PMNs and triplicate for remaining cells. The mean ± SEM for 24 and 48 h infected PMNs and 72 h time point for all other cell lines is displayed.

Figure S6.1



Morphologic differentiation with *A. phagocytophilum* infection. (A) Examples of neutrophils infected 25:1 MOI with *A. phagocytophilum*. At 3 and 48 h post infection, PMNs were stained with Wright-Giemsa stain and visualized at 1000X original for presence of infection. (B) Examples of non-differentiated and differentiated cells infected 25:1 MOI with *A. phagocytophilum*. At 3, 24 and 72 h post infection, cells were stained with Wright-Giemsa and visualized at 1000X for presence of infection. 72 h images are displayed. See text for *A. phagocytophilum* quantitation by qPCR. HSCs – hematopoietic stem cells.

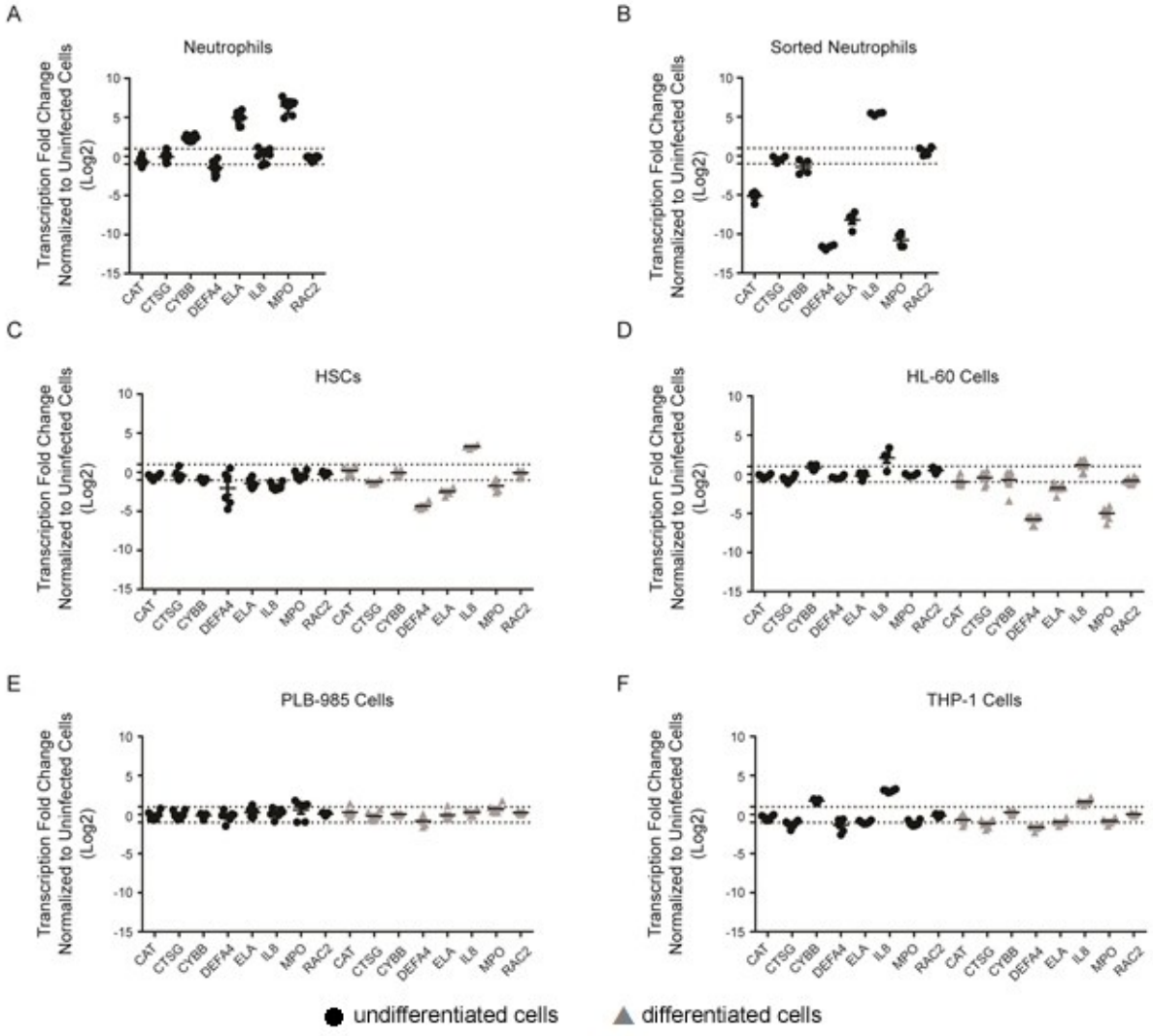
Transcription of cellular defense genes is altered with infection

A. phagocytophilum infection of granulocytes and neutrophils results in differential downregulation and upregulation of host genes. A number of defense genes are downregulated in *A. phagocytophilum*-infected THP-1 and HL-60 cells [13,14] and neutrophils. [15,22] In addition, chemokine gene expression is increased, and we used *IL8* as a monitor for upregulated gene expression. A panel of these genes was analyzed for expression in infected neutrophils, HSCs and the 3 cell lines (Fig. 6.2) to determine similarities of responses in cell lines vs. infected neutrophils. After fluorescence sorting to examine homogeneously-infected neutrophils, the defense gene expression profile was different than observed in the unsorted neutrophils, but consistent with that observed with differentiated HL-60 and HSCs (Fig. 6.2 B-D). Unsorted, heterogeneously-infected neutrophils, as observed in other studies, showed increased expression of *CYBB* ($gp91^{phox}$), *ELA* (elastase) and *MPO* (myeloperoxidase) (Fig. 6.2A), consistent with cellular activation of host defense pathways. In contrast, all 7 defense genes downregulated in granulocyte cell models of *A. phagocytophilum* infection were found repressed in the sorted homogeneously-infected population of neutrophils (Fig. 6.2B).

Of the cell lines, differentiated HL-60 cells, which were the most heavily and homogeneously infected without sorting, had the greatest number of defense genes downregulated (6 of 7) (Fig. 6.2D). Differentiated HSCs showed a similar pattern to HL-60 cells and neutrophils (Fig. 6.2C). Interestingly, not all genes previously shown to be downregulated in THP-1 cells had decreased transcript levels (Fig. 6.2D), possibly due to heterogeneity of infection or variations in transcriptional responses that accrued with continued passage of this cell line. For this study, we selected examination at 72 h post infection for all cells lines and not when >95% of cells were infected, which could have

biased toward HL-60 cells that are typically used for propagation, despite the use of low passage bacteria. Undifferentiated and differentiated PLB-985 cells had no significant changes in gene expression with infection (Fig. 6.2E), in accordance with the low or absent bacterial loads.

Figure 6.2.



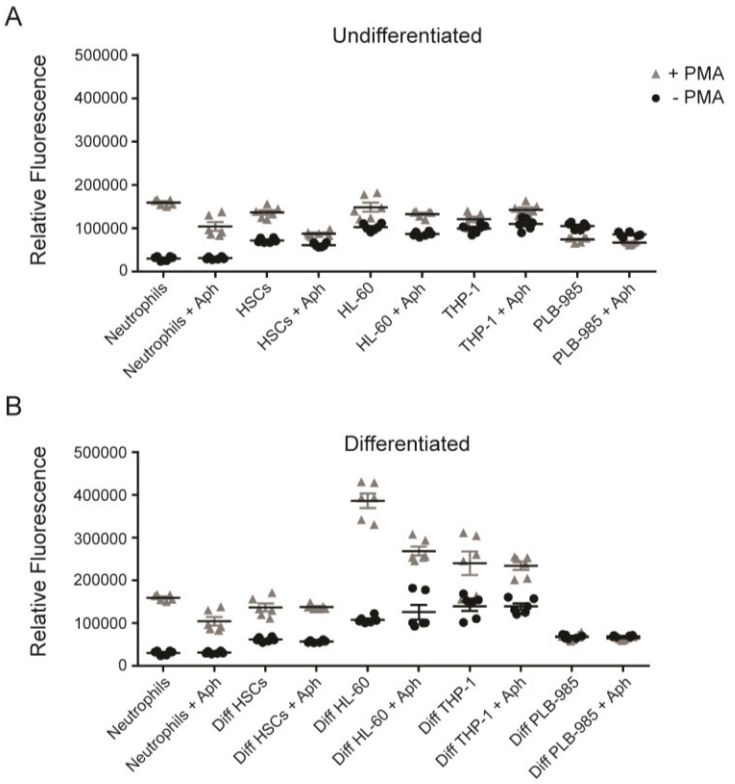
Defense gene transcription is altered with infection. Transcription of defense genes *CAT*, *CTSG*, *CYBB*, *DEFA4*, *ELA2*, *IL8*, *MPO*, and *RAC2* was assessed in infected cells relative to uninfected cells (fold change shown on y-axis). Infected PMN (A) and sorted mCherry-expressing *A. phagocytophilum*-infected PMNs (B) were normalized to the reference genes *B2M*, *GNB2LI*, *SDHA* and fold change in transcription at 24 h (●) post-infection was determined relative to 3 h uninfected PMNs (since uninfected neutrophils are apoptotic at 24 h *ex vivo*). HSCs (C) were normalized using *GNB2LI* and *RPL32* as

reference genes. HL-60 cells (D) were normalized with *GNB2LI* and *RPL13A*. PLB-985 cells (E) were normalized with the reference genes HMBS and *HPRT1*. THP-1 cells (F) were normalized using *ACTB* and *SDHA*. 72 h time points are displayed for the HSC, HL-60, PLB-985 and THP-1 cells for both undifferentiated (●) and differentiated (▲) cells. Infections were repeated in triplicate with technical duplicate readings for RT-PCR. Mean ± SEM is plotted. The dotted lines delimit the upper and lower 2-fold changes in gene expression.

A. phagocytophilum decreases intracellular oxidative burst

Suppression of superoxide production is critical for *A. phagocytophilum* survival in HL-60 cells after the first 3-5 hours of infection, [7,14] and is a hallmark of infected human and murine neutrophils *in vivo.* [11] Intracellular levels of superoxide anion were measured in response to the strong stimulant PMA in the presence and absence of *A. phagocytophilum* (Fig. 6.3). With differentiation, all cells increased production of superoxide anion in response to PMA except PLB-985 cells (Fig. 6.3B). Neutrophils, HSCs, and undifferentiated and differentiated HL-60 cells all decreased production of superoxide anion when infected with *A. phagocytophilum*. Infection did not decrease superoxide anion generation in differentiated HSCs, or in undifferentiated and differentiated PLB-985 and THP-1 cells (Fig. 6.3).

Figure 6.3



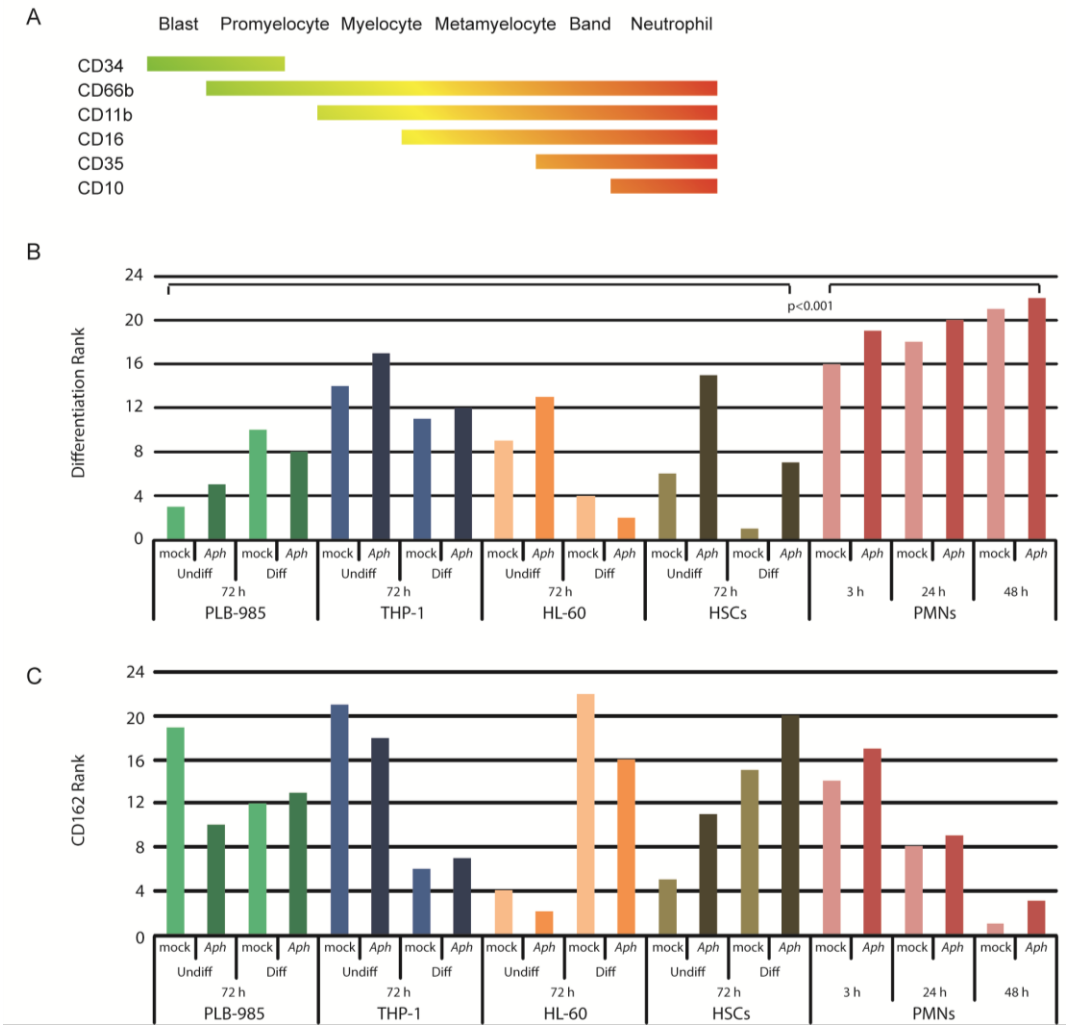
Granulocyte oxidative burst is decreased with *A. phagocytophilum* infection. Intracellular levels of reactive oxygen species in response to the stimulant PMA were detected using DCFH-DA 72 h post-infection for cell lines and HSCs and 24 h post-infection for neutrophils. The relative fluorescence at the 2 h time point was plotted for undifferentiated (A) and differentiated (B) cells with (*Aph)* and without infection. Infections were repeated in triplicate and each sample was read in duplicate. Mean ± SEM is plotted.

A. phagocytophilum infection does not affect cellular differentiation

Because the functions of *A. phagocytophilum*-infected neutrophils resemble those of less differentiated myeloid cells (e.g. decreased i] oxidative burst, ii] apoptosis, iii] margination, iv] emigration, and v] phagocytosis), [20,21,35,36] we examined the degree and effect of infection on differentiation by immunophenotype (Fig. 6.4B). As expected, neutrophils were most differentiated and varied considerably from all cell lines tested ($p<0.001$; Mann-Whitney test). *A. phagocytophilum*-infected cells did not show a significant change in differentiation compared to uninfected cells ($p=0.27$) suggesting that the bacterium does not target cellular differentiation or de-differentiation programs but is directed at specific host cell programs that impact functional changes. However, of 11 pairs of infected and uninfected cells, 9 infected samples were more differentiated than their uninfected counterparts. Of the infected/uninfected pairs, HSCs were markedly more differentiated when infected with *A. phagocytophilum* and only undifferentiated PLB-985 and undifferentiated HL-60 cells appeared less differentiated after infection.

A. phagocytophilum binds to CD162 to facilitate uptake [28,31] and infection-induced degranulation leads to shedding of surface CD162. [37] Differentiated HL-60 and HSC cells had the highest proportional increases in cells expressing CD162 with differentiation and HL-60 cells also demonstrated modest CD162 loss with infection (Fig. 4C). However, no overall significant differences in CD162 expression were detected when comparing infected and uninfected cells or neutrophils ($p=0.224$; Wilcoxon test). As anticipated, *ex vivo* neutrophils rapidly lost surface CD162 expression over 48 h, greater for uninfected cells probably because of apoptosis.

Figure 6.4



***A. phagocytophilum* alters cellular immunophenotype.** (A) Schematic diagram of the range over which granulocyte immunophenotypic surface markers are expressed with differentiation. In this panel, green represents less differentiation and red represents greater differentiation. (B) The expression of selected cellular markers (CD34, CD66b, CD11b, CD16, CD35 and CD10) of myeloid differentiation were measured using flow cytometry to determine differentiation state of uninfected and infected cellular models relative to *ex vivo* peripheral blood human neutrophils. The geometric mean of three

independent experiments was averaged for each immunophenotypic marker. The total fluorescence for a marker within the individual cell lines was calculated and used to determine the proportional expression of each marker for a particular cell line. The proportional expression was then weighted (-1 for CD34 , +1 for CD66b, +2 for CD11b, +3 for CD16, +4 CD35, and +5 for CD10). The overall differentiation status for each cell line and condition was determined by the total of the weighted scores for each marker. Undiff = undifferentiated, Diff = differentiated; Aph – *A. phagocytophilum* infected, mock – mock infected. The weighted scores were ranked and the Wilcoxon Rank-Sum test performed. (C) The CD162 geometric mean fluorescence of three independent experiments was averaged for each condition (cell line, differentiation status and time point). The proportion of CD162 expression was calculated within a cell line and ranked. Green bars – PLB-985 cells; blue bars – THP-1 cells; orange bars – HL-60 cells; brown bars – HSCs; pink/red bars – neutrophils (PMNs or polymorphonuclear neutrophils).

Discussion

A. phagocytophilum is a unique bacterium, in part because of its small genome and its obligate intracellular niche within the chief mammalian host defense cell, the neutrophil. In addition, mounting evidence shows that *A. phagocytophilum* evolved multiple mechanisms to subvert host cell functions as fitness advantages for survival and transmission, including the delivery of type IV secretion system effectors that not only mediate direct protein-protein disturbances of signaling but also bind host DNA, alter chromatin structure, and modify transcriptional programs governing host cell functions. [13,38,39] Advances in understanding the mechanisms by which *A. phagocytophilum* alters functional programs via signaling or transcriptional regulation is best accomplished using their natural mammalian target, neutrophils. Since experimental manipulation of neutrophils *ex vivo* is currently challenging, cell models that closely parallel neutrophil functions and that are validated to share underlying transcriptional and signaling programs would be beneficial. Currently, some accepted transcriptional responses proven important for *A. phagocytophilum* survival and transmission in cell lines have not been replicated in neutrophil transcriptome studies. [15,17] Because of these discrepancies, we sought to compare several existing and potentially useful granulocyte cell line models to determine which most closely mimics the functional, phenotypic, and transcriptional changes observed in neutrophils infected with *A. phagocytophilum*.

Higher levels of infection in all cell models including neutrophils, was clearly the major determinant of transcriptional and functional changes observed with infection. This was especially evident in differentiated HL-60 cells which were the most heavily and

homogeneously infected and concurrently had the most dramatic reductions in defense gene transcription and oxidative burst. This is in agreement with our prior studies and the results of other investigators. [7,13,16] In contrast, PLB-985 and THP-1 cells had low levels of infection and minimal change in function and defense gene expression, while NB4 cells never became sufficiently infected and were not further studied. In concert with these observations, the infection level in neutrophils also clearly influenced transcriptional changes as we found defense gene transcription was dramatically different in a homogeneously-infected neutrophil population with an average of >40 bacteria/cell as compared to unsorted heterogeneously-infected group with <3 bacteria/cell on average. In the heterogeneous neutrophils populations, the response of the more highly infected neutrophils was masked by the much larger population of uninfected neutrophils precluding the ability to observe the true response of *ex vivo* neutrophils to *A. phagocytophilum*. This observation in particular has important implications for studies of *A. phagocytophilum* conducted by infecting peripheral blood neutrophils *ex vivo*, when responses are analyzed within 12 or 24 hours post-infection. Hypothetically, the data suggest a higher intracellular bacterial load would result in the delivery of more effector molecules into the host cell to impact specific signaling pathways or transcriptional programs governing neutrophil functions.

Among the *A. phagocytophilum* infected cell models with homogeneous and high level infections, the most dramatic functional changes in antimicrobial responses, such as the lack of superoxide generation, are phenotypic features of immature progenitor cells. Given the relationship between cellular bacterial load and transcriptional and functional alterations, we investigated the effect of differentiation on bacterial load, transcription,

and respiratory burst for each cell line and HSCs to determine if *A. phagocytophilum* de-differentiates cells or reprograms specific cellular functions. The cellular immunophenotypes selected generally reflect maturation of neutrophil antimicrobial response and microbial detection. The findings here suggest that infection is either neutral or accelerates differentiation and maturation in cell lines and neutrophils. Surprisingly, PLB-985 cells were the least differentiated, and in addition to low bacterial load, this could also explain the absence of transcriptional and functional changes with infection. Furthermore, differentiated HL-60 and HSCs, like neutrophils, had the highest proportions of surface CD162, a key granulocyte adhesion molecule for *A. phagocytophilum* binding and invasion, [28,31] suggesting that its sustained expression allows a higher proportion of cells to become infected. Thus, the mechanism(s) that *A. phagocytophilum* uses to reprogram granulocyte function are not analogous to de-differentiation as might be observed with reprogramming of differentiated cells into induced pluripotent stem cells (iPSCs). Instead, these data indicate neutrophil/granulocyte reprogramming that results in differentially-regulated gene expression patterns evolved to target key cellular functions that improve microbial survival, propagation, and transmission.

A single cell *in vivo* that loses antimicrobial activity is unlikely to have much impact on disease. However, the loss of antimicrobial activity and the delayed apoptosis that occur with *A. phagocytophilum* could permit the infected cell to survive for several rounds of bacterial replication before disseminating to other susceptible granulocytes, which could be acquired in a subsequent tick blood meal. The proinflammatory response, including IL-8 and other upregulated chemokine/cytokine genes, clearly benefits bacterial

population expansion by recruiting new host cells.[6] Along with neutrophil degranulation, the proinflammatory response could have significant ancillary impacts on tissue injury and disease since these responses are amplified locally, as seen with the inflammatory lesions in tissues of mice, horses and humans.[40] In fact, disease with *A. phagocytophilum* infection is more closely linked to the overall host inflammatory and immune response to the bacterium than to the direct impact of infection in any one given neutrophil.[41,42] The cumulative effect whereby a median of 1% of circulating human neutrophils are infected, or about 1.7 x 10^8cells (1.7 x 10^9 bacteria) in a 70 kg human, could be dramatic.

How *A. phagocytophilum* alters antimicrobial responses, promotes self-survival and replication to be transmitted from reservoir hosts and to ultimately cause disease in humans will facilitate an improved understanding of microbial control of infected host cells, mechanisms to reduce disease severity, or could expand knowledge of how to control or contain undesired granulocyte responses in inflammatory diseases. Until better tractability of *ex vivo* neutrophils is achieved, susceptible, flexible, robust cell models that closely mimic neutrophils in function and transcriptional program responses are needed. The data here provide firm evidence that *A. phagocytophilum* infection studies designed to interrogate cellular functions in *ex vivo* neutrophils are flawed without approaches that achieve homogeneous infection at levels similar to those observed *in vivo*. These data also demonstrate that the well-described HL-60 cell infection model mimics and provides many desirable attributes by comparison. Moreover, it is now clear that the complex functional changes that occur in neutrophils or granulocyte cells infected by *A. phagocytophilum* are not the result of a de-differentiation program as seen with iPSCs. Rather, these changes occur from targeted reprogramming of functions that

affect specific transcriptional and/or signaling pathways. This ultimately benefits microbial survival and transmission, but can concurrently generate inflammatory injury and disease.

References

1. Choi KS, Park JT, Dumler JS. *Anaplasma phagocytophilum* delay of neutrophil apoptosis through the p38 mitogen-activated protein kinase signal pathway. *Infect Immun*. 2005;73(12):8209-8218.

2. Ge Y, Rikihisa Y. *Anaplasma phagocytophilum* delays spontaneous human neutrophil apoptosis by modulation of multiple apoptotic pathways. *Cell Microbiol*. 2006;8(9):1406-1416.

3. Yoshiie K, Kim HY, Mott J, Rikihisa Y. Intracellular infection by the human granulocytic ehrlichiosis agent inhibits human neutrophil apoptosis. *Infect Immun*. 2000;68(3):1125-1133.

4. Akkoyunlu M, Malawista SE, Anguita J, Fikrig E. Exploitation of interleukin-8-induced neutrophil chemotaxis by the agent of human granulocytic ehrlichiosis. *Infect Immun*. 2001;69(9):5577-5588.

5. Klein MB, Hu S, Chao CC, Goodman JL. The agent of human granulocytic ehrlichiosis induces the production of myelosuppressing chemokines without induction of proinflammatory cytokines. *J Infect Dis*. 2000;182(1):200-205.

6. Scorpio DG, Akkoyunlu M, Fikrig E, Dumler JS. CXCR2 blockade influences *Anaplasma phagocytophilum* propagation but not histopathology in the mouse model of human granulocytic anaplasmosis. *Clin Diagn Lab Immunol*. 2004;11(5):963-968.

7. Banerjee R, Anguita J, Roos D, Fikrig E. Cutting edge: Infection by the agent of human granulocytic ehrlichiosis prevents the respiratory burst by down-regulating gp91phox. *J Immunol*. 2000;164(8):3946-3949.

8. IJdo JW, Carlson AC, Kennedy EL. *Anaplasma phagocytophilum* AnkA is tyrosine-phosphorylated at EPIYA motifs and recruits SHP-1 during early infection. *Cell Microbiol*. 2007;9(5):1284-1296.

9. Carlyon JA, Abdel-Latif D, Pypaert M, Lacy P, Fikrig E. *Anaplasma phagocytophilum* utilizes multiple host evasion mechanisms to thwart NADPH oxidase-mediated killing during neutrophil infection. *Infect Immun*. 2004;72(8):4772-4783.

10. Choi KS, Dumler JS. Early induction and late abrogation of respiratory burst in *A. phagocytophilum*-infected neutrophils. *Ann N Y Acad Sci*. 2003;990:488-493.

11. Wang T, Malawista SE, Pal U, et al. Superoxide anion production during *Anaplasma phagocytophila* infection. *J Infect Dis*. 2002;186(2):274-280.

12. Thomas V, Samanta S, Wu C, Berliner N, Fikrig E. *Anaplasma phagocytophilum* modulates gp91phox gene expression through altered interferon regulatory factor 1 and

PU.1 levels and binding of CCAAT displacement protein. *Infect Immun*. 2005;73(1):208-218.

13. Garcia-Garcia JC, Rennoll-Bankert KE, Pelly S, Milstone AM, Dumler JS. Silencing of host cell CYBB gene expression by the nuclear effector AnkA of the intracellular pathogen *Anaplasma phagocytophilum*. *Infect Immun*. 2009;77(6):2385-2391.

14. Carlyon JA, Chan WT, Galan J, Roos D, Fikrig E. Repression of *rac2* mRNA expression by *Anaplasma phagocytophila* is essential to the inhibition of superoxide production and bacterial proliferation. *J Immunol*. 2002;169(12):7009-7018.

15. Borjesson DL, Kobayashi SD, Whitney AR, Voyich JM, Argue CM, Deleo FR. Insights into pathogen immune evasion mechanisms: *Anaplasma phagocytophilum* fails to induce an apoptosis differentiation program in human neutrophils. *J Immunol*. 2005;174(10):6364-6372.

16. de la Fuente J, Ayoubi P, Blouin EF, Almazan C, Naranjo V, Kocan KM. Gene expression profiling of human promyelocytic cells in response to infection with *Anaplasma phagocytophilum*. *Cell Microbiol*. 2005;7(4):549-559.

17. Lee HC, Kioi M, Han J, Puri RK, Goodman JL. *Anaplasma phagocytophilum*-induced gene expression in both human neutrophils and HL-60 cells. *Genomics*. 2008;92(3):144-151.

18. Pedra JH, Sukumaran B, Carlyon JA, Berliner N, Fikrig E. Modulation of NB4 promyelocytic leukemic cell machinery by *Anaplasma phagocytophilum*. *Genomics*. 2005;86(3):365-377.

19. Sukumaran B, Carlyon JA, Cai JL, Berliner N, Fikrig E. Early transcriptional response of human neutrophils to *Anaplasma phagocytophilum* infection. *Infect Immun*. 2005;73(12):8089-8099.

20. Choi KS, Grab DJ, Dumler JS. *Anaplasma phagocytophilum* infection induces protracted neutrophil degranulation. *Infect Immun*. 2004;72(6):3680-3683.

21. Garyu JW, Choi KS, Grab DJ, Dumler JS. Defective phagocytosis in *Anaplasma phagocytophilum*-infected neutrophils. *Infect Immun*. 2005;73(2):1187-1190.

22. Lee HC, Kioi M, Han J, Puri RK, Goodman JL. *Anaplasma phagocytophilum*-induced gene expression in both human neutrophils and HL-60 cells. *Genomics*. 2008;92(3):144-151.

23. Breitman TR, Selonick SE, Collins SJ. Induction of differentiation of the human promyelocytic leukemia cell line (HL-60) by retinoic acid. *Proc Natl Acad Sci U S A*. 1980;77(5):2936-2940.

24. Drach J, Lopez-Berestein G, McQueen T, Andreeff M, Mehta K. Induction of differentiation in myeloid leukemia cell lines and acute promyelocytic leukemia cells by liposomal all-trans-retinoic acid. *Cancer Res*. 1993;53(9):2100-2104.

25. Idres N, Benoit G, Flexor MA, Lanotte M, Chabot GG. Granulocytic differentiation of human NB4 promyelocytic leukemia cells induced by all-trans retinoic acid metabolites. *Cancer Res*. 2001;61(2):700-705.

26. Tucker KA, Lilly MB, Heck L,Jr, Rado TA. Characterization of a new human diploid myeloid leukemia cell line (PLB-985) with granulocytic and monocytic differentiating capacity. *Blood*. 1987;70(2):372-378.

27. Davies RS, Madigan JE, Hodzic E, Borjesson DL, Dumler JS. Dexamethasone-induced cytokine changes associated with diminished disease severity in horses infected with anaplasma phagocytophilum. *Clin Vaccine Immunol*. 2011;18(11):1962-1968.

28. Goodman JL, Nelson C, Vitale B, et al. Direct cultivation of the causative agent of human granulocytic ehrlichiosis. *N Engl J Med*. 1996;334(4):209-215.

29. Felsheim RF, Herron MJ, Nelson CM, et al. Transformation of *Anaplasma phagocytophilum*. *BMC Biotechnol*. 2006;6:42.

30. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods*. 2001;25(4):402-408.

31. Herron MJ, Nelson CM, Larson J, Snapp KR, Kansas GS, Goodman JL. Intracellular parasitism by the human granulocytic ehrlichiosis bacterium through the P-selectin ligand, PSGL-1. *Science*. 2000;288(5471):1653-1656.

32. Popov VL, Han VC, Chen SM, et al. Ultrastructural differentiation of the genogroups in the genus *Ehrlichia*. *J Med Microbiol*. 1998;47(3):235-251.

33. Choi KS, Park JT, Dumler JS. *Anaplasma phagocytophilum* delay of neutrophil apoptosis through the p38 mitogen-activated protein kinase signal pathway. *Infect Immun*. 2005;73(12):8209-8218.

34. Lee HC, Goodman JL. *Anaplasma phagocytophilum* causes global induction of antiapoptosis in human neutrophils. *Genomics*. 2006;88(4):496-503.

35. Scaife H, Woldehiwet Z, Hart CA, Edwards SW. *Anaplasma phagocytophilum* reduces neutrophil apoptosis in vivo. *Infect Immun*. 2003;71(4):1995-2001.

36. Mott J, Rikihisa Y, Tsunawaki S. Effects of *Anaplasma phagocytophila* on NADPH oxidase components in human neutrophils and HL-60 cells. *Infect Immun*. 2002;70(3):1359-1366.

37. Choi KS, Garyu J, Park J, Dumler JS. Diminished adhesion of *Anaplasma phagocytophilum*-infected neutrophils to endothelial cells is associated with reduced expression of leukocyte surface selectin. *Infect Immun*. 2003;71(8):4586-4594.

38. Garcia-Garcia JC, Barat NC, Trembley SJ, Dumler JS. Epigenetic silencing of host cell defense genes enhances intracellular survival of the rickettsial pathogen *Anaplasma phagocytophilum*. *PLoS Pathog*. 2009;5(6):e1000488.

39. Lin M, den Dulk-Ras A, Hooykaas PJ, Rikihisa Y. *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by abl-1 to facilitate infection. *Cell Microbiol*. 2007;9(11):2644-2657.

40. Lepidi H, Bunnell JE, Martin ME, Madigan JE, Stuen S, Dumler JS. Comparative pathology, and immunohistology associated with clinical illness after *Ehrlichia phagocytophila*-group infections. *Am J Trop Med Hyg*. 2000;62(1):29-37.

41. Choi KS, Webb T, Oelke M, Scorpio DG, Dumler JS. Differential innate immune cell activation and proinflammatory response in *Anaplasma phagocytophilum* infection. *Infect Immun*. 2007;75(6):3124-3130.

42. Martin ME, Caspersen K, Dumler JS. Immunopathology and ehrlichial propagation are regulated by interferon-gamma and interleukin-10 in a murine model of human granulocytic ehrlichiosis. *Am J Pathol*. 2001;158(5):1881-1888.

Chapter 7: Conclusions and Future Directions

Portions of this work have been published as a review to *Frontiers in Genetics* and are reprinted here with permission from Frontiers Media.

Sinclair SH, Rennoll-Bankert KE, and JS Dumler. 2014. Effector bottleneck: microbial reprogramming of parasitized host cell transcription by epigenetic remodeling of chromatin structure. *Front Genet.* Aug 14; 5:274.

The ability of *Anaplasma phagocytophilum* replicate within human neutrophils requires the bacterial manipulation of host cellular processes, namely antimicrobial functions. To date, many studies have been dedicated to elucidating which programs of the neutrophil are altered during the course of infections and suggesting that many changes are regulated at the level of transcription. This thesis proposes and describes a mechanism for widespread changes observed in the neutrophil. The combination of large coordinately regulated chromosomal territories and hypermethylation of the infected cell genome suggest that *A. phagocytophilum* acts on a global scale to alter its host cell, not via one protein or individually targeted genes. Furthermore, bioinformatic and initial analysis implies that the bacterium encodes additional nucleomodulins beyond the well-studied AnkA. Together, these data begin to describe the complex interplay between parasite and host and suggests host modification requires relatively few effectors to inflict widespread alterations in host cell transcription and the neutrophil functions these govern.

Given the limited bacterial genome complexity compared with the expansive host genome complexity and its organization and complexity of transcriptional regulation in eukaryotes, the ability of prokaryotes to alter eukaryotic host cell functions at the epigenetic level is truly a remarkable phenomenon. The processes by which bacteria-derived proteins alter host cell signaling, endocytic or vesicular trafficking, and *cis* transcription of genes are important. However, these pathways become challenging as paradigms that can account for the marked and diverse changes in host cell transcription and functions during the course of infection. We believe that successful intracellular prokaryotes, whether parasitic and pathogenic, mutualistic, or symbiotic, evolved multifunctional complex proteins with broad effects that target key master regulators or

checkpoints for major cellular reprogramming events. The *A. phagocytophilum* genome is a fraction of the size of its human host, yet it manages to disrupt core functions of the cell by transforming the transcriptome and improving microbial fitness for survival and transmission. It is increasingly apparent that *A. phagocytophilum* and perhaps other intracellular prokaryotes manipulate their hosts with a high degree of efficacy, but alterations of the 'on' or 'off' state of genes in a *cis* only fashion or targeting individual or small numbers of signaling pathways is unlikely to account for this alone, in essence, an "effector bottleneck". Analysis of available infection transcriptome data demonstrates that large chromosomal territories appear to be coordinately regulated - up or down. This observation, along with widespread DNA methylation and HDAC recruitment by AnkA, lends strong evidence to the idea that *A. phagocytophilum* induces global changes in host gene expression via broad mechanisms that involve epigenetic regulation of transcriptional programs which belie cellular functions.

Elucidating the exact locations where AnkA or other nucleomodulins bind throughout the genome, and determining the architecture of the surrounding chromatin will be crucial to understanding this process. The interplay between HDACs, DNMTs and methyl binding proteins suggests that *A. phagocytophilum* infection could also induce widespread DNA methylation perhaps as a mechanism for obtaining broad epigenetic changes and functional reprogramming.

Understanding the role of prokaryotic control over complex eukaryotic transcription machinery in lieu of the bottleneck that would occur with single effector targets at signaling pathways will allow for better understanding of the essential components of whole cell transcriptional re-programming. However, this has

implications not only for host-pathogen or host-symbiont interactions, but across biology. The ability of proteins like AnkA of *A. phagocytophilum* and OspF of *S. flexneri* to alter inflammatory responses could provide opportunities for engineering of therapeutic agents which interfere with cellular responses in inflammatory diseases or other conditions where epigenetic factors control or contribute to disease.

Curriculum Vitae

Sara Hideko Sinclair

PO Box 1962 Chinle, AZ 86503 • (208) 250-7436 • gilm5447@gmail.com

Work Experience:

Dine College: Tsaile, AZ **2014-current**
Science Division
Instructor for Microbiology, Anatomy and Physiology

Education:

Johns Hopkins University: Baltimore, MD **2009-2014**
Department of Cellular and Molecular Medicine
PhD Candidate – estimated graduation December 2014
GPA not officially assigned, ~3.5/4.0

Idaho State University: Pocatello, ID **2008-2009**
College of Pharmacy
Masters in Pharmacology and Pharmaceutical Sciences
GPA – 4.0/4.0

University of Idaho – Moscow, ID **2004-2007**
College of Agricultural and Life Sciences
Bachelor of Science in Molecular Biology and Biochemistry
GPA – 3.79/4.0

Research Experience:

Johns Hopkins University, School of Medicine
Department of Cellular and Molecular Medicine
PhD Candidate (Aug 2009 – Aug 2014)

- Investigating the ability of the pathogen *Anaplasma phagocytophilum* to induce transcriptional changes in its host cell, the neutrophil, through epigenetic modifications.
- Developed an assay for identifying novel nuclear bacterial effector proteins
- Used TIBCO Spotfire and other bioinformatics tools to analyze next generation sequencing data sets

- Mentored and aided in the training of two technicians and five visiting scientists to become more independent scientists
- Attended multiple scientific conferences to discuss my research to a wide audience of scientists both within the rickettsial and microbiology communities.

Idaho State University
Department of Biomedical and Pharmaceutical Science
Graduate Research Assistantship (Aug 2008 – July 2009)

- Measured activation of the stress response to alcohol in cycling female rats by quantifying the amount of corticotropin- releasing hormone secreted by the paraventricular nucleus of the hypothalamus in rats in each of the four stages of estrus.
- Coordinated a team of four researchers to complete complex animal studies including two different animal surgeries and intensive schedules of hormone injections.

University of Idaho
Department of Microbiology, Molecular Biology, and Biochemistry
Undergraduate Research Assistant – Dr. Scott Kobayashi (Sept 2006 – Dec 2007)

- Created a library of gene knockouts in *Yersinia pestis* strains KIM5 and KIM6+.
- Characterizing the interactions between human neutrophils and *Y. pestis* LPS.
- *Undergraduate Research Assistant – Dr. Gary Daughdrill (Jan 05 – Sept 06)*
- Performed basic lab techniques necessary for purifying proteins using FPLC (Fast Protein Liquid Chromatography).
- Analyzed NMR (Nuclear Magnetic Resonance) data to detect the intrinsic structure of flexible protein linkers.

Publications, Posters, and Presentations:

Publications

- **Sinclair SH**, Yegnasubramanian S, Dumler JS. (2015). Global DNA methylation changes and differential gene expression in *Anaplasma phagocytophilum*-infected human neutrophils. *Submitted to Genome Biology.*
- **Sinclair SH**, Garcia-Garcia JC, Dumler JS. (2015). Bioinformatic and mass spectrometry identification of *Anaplasma phagocytophilum* proteins translocated into host cell nuclei. *Front Microbio.* 6(55):1-10
- **Sinclair SH**, Rennoll-Bankert KE, Dumler JS (2014). Effector bottleneck: microbial reprogramming of parasitized host cell transcription by epigenetic remodeling of chromatin structure. *Front Genet.* Aug 14; 5:274 PMID: 25177343
- Rennoll-Bankert KE, Garcia Garcia JC, **Sinclair SH**, Dumler JS. (2014). Chromatin bound bacterial effector AnkA recruits HDAC1 and modifies host genome expression. *Submitted To Cellular Microbiology – Under Revision.*
- Rennoll-Bankert KE, **Sinclair SH**, Lichay MA, Dumler JS (2014). Comparison and characterization of granulocyte cell models for *Anaplasma phagocytophilum* infection. *Pathog Dis.* 71(1):55-64 PMID:24376092

• Daughdrill GW, Narayanaswami P, **Gilmore SH**, Belczyk A, Brown CJ (2007). Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino Acid sequence conservation. *J Mol Evol*. 65, 277- 88. PMID: 17721672

Talks and Presentations

• **Gilmore SH**, Rennoll-Bankert KE, Dumler JS (2011). Microbial manipulation of host transcriptional programming by nuclear toxins – the *Anamplasma phagocytophilum* AnkA model. Presented at the 3rd International Young Researches Seminar, Sapporo, Hokkaido, Japan.

Posters and Abstracts

• **Sinclair SH**, Dumler JS(2013). *Anaplasma phagocytophilum* infection increases genome-wide DNA methylation in human neutrophils. Presented at the American Society of Rickettsiology meeting. Portland, ME, USA.

• **Gilmore SH**, Garcia-Garcia JC, Dumler JS (2012). The identification of 6 novel nuclear proteins in *Anaplasma phagocytophilum*. Presented at the Mid Atlantic Microbial Pathogenesis Meeting, American Society of Microbiology and meetings. Roseland, VA and Denver, CO, USA.

• Rennoll-Bankert KE, **Gilmore SH**, Lichay MA, Dumler JS (2012). Comparison and characterization of current granulocyte cell models for *Anaplasma phagocytophilum* infection. Presented at American Society of Rickettsiology. Park City, Utah, USA.

• **Gilmore SH**, Rennoll-Bankert KE, Dumler JS (2011). Microbial manipulation of host transcriptional programming by nuclear toxins – the *Anaplasma phagocytophilum* AnkA model. Presented at the 3rd International Young Researches Seminar. Sapporo, Hokkaido, Japan.

• **Gilmore SH**, Dumler JS (2011). Infection with *Anaplasma phagocytophilum* alters specific transcriptional programs but does not appear to induce cellular de-differentiation. Presented at Rickettsia2011. Heraklion, Crete, Greece.

• **Gilmore SH**, Dumler JS, Garcia-Garcia JC (2010). DNA methylation and epigenetic regulation of granulocyte antimicrobial and cell cycle genes with *Anaplasma phagocytophilum* infection. Presented at American Society of Rickettsiology. Portland, OR, USA.

• **Gilmore SH**, Binks SL, Larkin JW, Li Y, and Selvage DJ (2008). Effects of estradiol on HPA axis responses to icv alcohol in the rat. Presented at the Snake River Association of Neuroscience annual conference. Pocatello, ID, USA.

• Larkin JW, Li Y, Binks SL, **Gilmore SH**, and Selvage DJ (2008). Roles of the locus coeruleus and norepinephrine in the activation of the HPA axis induced via central alcohol administration in the rat. Snake River Association for Neuroscience annual

conference. Pocatello, ID, USA.

- Gower J, Larkin JW, Binks SL, **Gilmore SH**, and Selvage DJ (2008). Role Of Norepinephrine In HPA Activation By Intraventricular Ethanol. The annual Idaho IDeA Network of Biomedical Research Excellence (INBRE) conference. Boise, ID, USA.
- **Gilmore SH** and Kobayashi SD (2007). Effects of *Yersinia pestis* lipid A acylation o neutrophil priming. Presented at the annual Idaho IDeA Network of Biomedical Researcy Excellence annual conference. Moscow, ID, USA.
- **Gilmore SH**, O'Loughli L, and Kobayashi SD (2007). The role of two-component systems in *Yersinia pestis* interactions with human neutrophils. Presented at the annual University of Idaho Undergraduate Research Symposium. Moscow, ID, USA.
- **Gilmore SH**, Narayanaswami P, Stancik A, Lowry D, Shrestha R, Vise P, and Daughdrill GW (2006). Determining the average structure of intrinsically unstructured linker domains from the 70kDa subunit of replication protein A using nuclear magnetic resonance spectroscopy. Presented at the annual Idaho IDeA Network of Biomedical Research Excellence conference. Coeur d'Alene, ID, USA.
- **Gilmore SH**, Narayanaswami P, Stancik A, Vise P, and Daughdrill GW (2006). Determining the average structure of the intrinsically unstructured linker domains of replication protein A in four species using nuclear magnetic spectroscopy. Presented at the annual University of Idaho Undergraduate Research Symposium. Moscow, ID, USA.

Teaching Experience:

Teaching Assistant for the department of Cellular and Molecular Medicine, Molecular Biology and Genomics (2011)

- Ran weekly review sessions for departmental students and provided one-on-one tutoring to students in need for the Molecular Biology and Genomics course.

Teaching Assistant for Dr. Doug Cole, Intro. to Biochemistry (2007)

- Graded weekly homework assignments and held 6 office hours/week.

Activities:

Lead Mentor and Team Lead • Incentive Mentoring Program (2009-2014)

- Managed and trained four Team Leads to become effective leaders ensuring students' academic and social development
- Coordinated teams of 8-10 mentors who provide 'family' style support to a student by providing tutoring, community service projects, college prep and social activities
- Two of my students are currently attending college and two are on track for graduating high school

PetValu Lead • Baltimore Animal Rescue and Care Shelter (2012-2013)

- Managed 16 volunteers at an offsite location and aided in adoption and placement

of 40+ cats this year
- Success of this offsite adoption center led to the opening of another center at a neighboring PetValu store in 2013

Entrepreneur: I Felt Fuzzy & Syringa Paper Co (2011-current)
- Started my own small businesses using Etsy.com as an e-marketplace platform to sell my artwork and paper crafts
- Performed all photography, marketing, social media campaigns, finances, and product development to date
- Sold 140+ pieces online, and designed a website to expand web presence, http://ifeltfuzzy.com
- Featured in the 'New!' section of craft magazine *Mollie Makes* for embroidery hoop art

Associated Graduate Students of Pharmacy (2008 – 2009)

Graduate Student Association and Interclub Council Representative (Fall 2009)

College of Agricultural & Life Sciences (CALS) Ambassador (April 2006 – Dec 2007)
- Travel to high schools across the Pacific Northwest to promote the university and CALS.

Microbiology, Molecular Biology, and Biochemistry Departmental Advisory Board (2006 & 2007)

President and Secretary • Microbiology, Molecular Biology, and Biochemistry Club (Sept 2004 – Dec 2007)
- President (May 2006 – May 2007) & Secretary (May 2005 – May 2006)
- Organized community outreach programs to teach germ awareness at three local elementary schools

College of Agricultural & Life Sciences Peer Advisor (April 2006 – Dec 2006)
- Assisted incoming students with adjusting to college and advising on course loads

Gamma Sigma Delta, (inducted May 2007) - Agricultural Honor Society

Honors and Awards:
- Incentive Mentoring Program IMPact Award – 2014
- Governor's Service Award – 2013 • Incentive Mentoring Program Leadership Award - 2012
- Research Assistantship – Academic year 2008-2009
- Idaho IDeA Network of Biomedical Research Excellence Student of the Year – 2007
- Idaho IDeA Network of Biomedical Research Excellence Summer Research Fellowship – Summers of 2006 and 2007 • University of Idaho Undergraduate Research Symposium Poster Contest, 1st place – Spring 2007
- Gamma Sigma Delta - Agricultural Honor Society – inducted May 2007 • University of Idaho Undergraduate Research Symposium Poster Contest, 2nd place – Spring 2006
- Microbiology, Molecular Biology, and Biochemistry Senior of the Year – Spring 2007
- Microbiology, Molecular Biology, and Biochemistry Freshman of the Year - Spring 2005

- Idaho Space Grant Consortium NASA Scholar, 2004-2007
- James Anderson and Kay Bozar Scholarship – Fall 2007 • Taylor Mary Branson Scholarship – Fall 2007
- JR Simplot Company Scholarship – 2006 and 2007
- Deans List, Dec. 2004 – Dec. 2007
- University of Idaho Top Scholar – Spring 2004

References:

Dr. J Stephen Dumler, PhD
University of Maryland, School of Medicine – Departments of Pathology and Microiology
Johns Hopkins University, School of Medicine – Department of Pathology
410-706-5072
sdumler@som.umaryland.edu

Dr. Dan Selvage, PhD
University of New England – College of Pharmacy
207-221-4059
dselvage@une.edu

Dr. Scott Kobayashi, PhD
National Institute of Allergy and Infectious Disease – Laboratory of Human Bacterial Pathogenesis
406-375-9704
kobayashis@niaid.nih.gov

Dr. Gary Daughdrill, PhD
University of Southern Florida – Department of Cell Biology, Microbiology and Molecular Biology
813-974-2503
gdaugh@uidaho.edu